

WESTFÄLISCHE  
WILHELMS-UNIVERSITÄT  
MÜNSTER

# PET-BILDREKONSTRUKTION IM PHASENRAUM

MASTERARBEIT

Westfälische Wilhelms-Universität Münster  
Fachbereich Mathematik und Informatik  
Institut für Angewandte Mathematik

Eingereicht von: Matthias Fröhlich  
Datum: 26. August 2021  
Erstkorrektur: Prof. Dr. Benedikt Wirth  
Zweitkorrektur: Dr. Frank Wübbeling



# Inhaltsverzeichnis

<b>1</b>	<b>Vorwort</b>	<b>1</b>
<b>2</b>	<b>Grundlagen</b>	<b>5</b>
2.1	Konzepte der Maßtheorie und Funktionalanalysis . . . . .	5
2.2	Direkte Methode . . . . .	15
2.3	Stochastische Modellierung . . . . .	19
2.4	Poisson-Punkt-Prozesse . . . . .	22
<b>3</b>	<b>Der Messoperator auf dem Phasenraum</b>	<b>25</b>
3.1	Funktionsweise der Positronen-Emissions-Tomographie . . . . .	25
3.2	Der Vorwärtsoperator . . . . .	26
3.3	Dynamische Super-Resolution . . . . .	30
3.4	Der dynamische Messoperator . . . . .	34
<b>4</b>	<b>Maximum-aposteriori-Schätzung und das Rekonstruktionsfunktional</b>	<b>41</b>
4.1	Eine erste Variante des Rekonstruktionsfunktionals . . . . .	41
4.2	Problematik und Unterscheidung der Daten . . . . .	43
4.3	Gewichtetes Maximums-Funktional . . . . .	45
4.4	Konvex relaxiertes Rekonstruktionsfunktional . . . . .	48
<b>5</b>	<b>Existenz von Minimierern</b>	<b>51</b>
<b>6</b>	<b>Numerische Simulation</b>	<b>61</b>
6.1	Frank-Wolfe-Algorithmus . . . . .	61
6.2	Diskretisierung und die Richtungsableitung . . . . .	63
6.3	Die Implementierung und Simulation der Messungen . . . . .	68
6.4	Numerische Ergebnisse . . . . .	70
<b>7</b>	<b>Ausblick</b>	<b>73</b>
	<b>Literaturverzeichnis</b>	<b>74</b>



# 1 Vorwort

In der vorliegenden Arbeit beschäftigen wir uns mit einem Bildrekonstruktionsverfahren, das Anwendung in der Positronen-Emissions-Tomographie (PET) findet. Im Detail wird eine Methode untersucht, die den Rückschluss auf Position und Geschwindigkeit sich linear bewegender Teilchen erlaubt. Die PET ist ein medizinisches Bildgebendes Verfahren, bei dem das Untersuchungsobjekt eine geringe Menge an radioaktiven Isotopen verabreicht bekommt. Die zufällige Strahlung wird indirekt von den Detektoren eines Tomographen gemessen und als Datensatz gespeichert. Mathematisch werden die Isotope durch ein Maß modelliert, dessen Form einer Linearkombination von Dirac-Maßen

$$\mu = \sum_{i=1}^N w_i \delta_{x_i + tv_i}$$

entspricht. Hier werden insgesamt  $N$ -Isotope beschrieben mit Positionen  $x_i \in \Sigma$  innerhalb des Tomographen  $\Sigma \subset \mathbb{R}^d$  und zugehörigen Gewichten  $w_i \in \mathbb{R}$ . Zu jedem der Teilchen gehört außerdem ein Geschwindigkeitsvektor  $v_i \in \mathbb{R}^d$ . Der Definition von  $\mu$  entsprechend bewegen sich die Isotope während eines Messzeitraums  $[0, T]$  linear, wobei  $T > 0$  und  $t \in [0, T]$  gilt. Das Ziel eines jeden Rekonstruktionsverfahrens besteht folglich aus dem Ermitteln von  $\mu$  mittels gegebener Messungen. In ihrem Artikel „Towards a Mathematical Theory of Super-resolution“ [7] beschreiben die Autoren Candès und Fernandez-Granda eine Möglichkeit, wie eine Rekonstruktion vonstatten gehen kann. Für einen in unserer Anwendung festen Zeitpunkt  $t \in [0, T]$  betrachten sie das Minimierungsproblem

$$\min_{\lambda \in M(\Sigma)} \|\lambda\|_{TV} \quad , \text{ so dass } E(\lambda) = \tilde{y}$$

und beweisen, dass sich jede Position der  $N$  Isotope,  $\{(x_i + tv_i) \in \Sigma \mid i = 1, \dots, N\}$ , durch dessen Lösung rekonstruieren lässt, soweit die Isotope einen Mindestabstand zueinander einhalten. Dabei beschreibt der Operator  $E$  den Messvorgang und  $\tilde{y} = E(\mu)$  die gemessenen Daten. Als Norm, über die minimiert werden soll, wählen die Autoren die Totale-Variations-Norm. Durch wiederholtes Messen, so dass insgesamt  $K \in \mathbb{N}$  Messungen vollzogen werden, kann neben den Positionen auch auf die Geschwindigkeiten der Isotope geschlossen werden. Indem das Minimierungsproblem nach Candès et al. zu mehreren Zeitpunkten  $t_k \in [0, T]$ ,  $k \in \{0, \dots, K-1\}$ , aufgestellt und gelöst wird, erhalten wir  $K$  verschiedene Positionsdaten. Mittels anschließender Interpolation über die Positionen und mit dem Wissen über den Zeitunterschied zwischen zwei Messungen,  $t_k - t_{k-1} = \frac{T}{K}$ , lassen sich die Geschwindigkeiten der Isotope bestimmen.

In dieser Arbeit analysieren wir eine alternative Rekonstruktionsmethode, erstmals erwähnt

von Alberti et al. in ihrem Paper „Dynamic Spike Super-resolution and Applications to Ultrafast Ultrasound Imaging“ [9]. Anstelle der einzelnen Positions-Rekonstruktion aller  $K$  Messungen beschreiben Alberti et al. ein Vorgehen, wie Positionen und Geschwindigkeiten simultan ermittelt werden können. Dabei werden Maße auf dem sogenannten *Phasenraum*  $\Omega$  betrachtet. Dieser verknüpft sämtliche Ortsvektoren von  $\Sigma$  mit möglichen Geschwindigkeitsvektoren aus dem  $\mathbb{R}^d$ . Die Struktur der betrachteten Maße ist der Struktur von  $\mu$  ähnlich und gegeben durch eine Linearkombination von Dirac-Maßen auf  $\Omega$

$$\omega = \sum_{i=1}^N w_i \delta_{(x_i, v_i)} ,$$

wobei hier  $(x_i, v_i) \in \Omega$  das zum  $i$ -ten Isotop gehörende Paar von Position  $x_i$  und Geschwindigkeit  $v_i$  definiert. Durch das Aufstellen und Lösen eines vorherigem entsprechenden Minimierungsproblems

$$\min_{\lambda \in M(\Omega)} \|\lambda\|_{TV} \quad , \text{ so dass } G(\lambda) = \hat{y} \quad , \quad (1.1)$$

sind wir nach Candès et al. in der Lage, alle Elemente (Tupel) der sogenannten *Teilchenkonfiguration*  $\{(x_i, v_i) \in \Omega \mid i = 1, \dots, N\}$  zu rekonstruieren. Wie zuvor, beschreibt der Operator  $G$  hierbei den Messprozess, ebenso sind die Daten  $\hat{y}$  erneut von der Form  $\hat{y} = G(\omega)$ .

Neben den Vorteilen, die diese Methode verspricht und auf die im Abschnitt 3.3 näher eingegangen wird, birgt die Betrachtung von Maßen auf  $\Omega$  auch einen hervorzuhebenden Nachteil. Durch die Kombination von Positionen und Geschwindigkeiten als Tupel besitzt das Minimierungsproblem Dimension  $2d$ . Bei drei Raumdimensionen ist demnach für eine numerische Lösung ein sechsdimensionales Gitter nötig. Damit gehen hohe Anforderungen an die Rechenkapazität von Computern einher, die zu langen Laufzeiten des implementierten Lösungsalgorithmus führen. Abhilfe für dieses Problem verschaffen Dimensionsreduzierende Verfahren, wie zum Beispiel von Alexander Schlueter in seiner Masterthesis vorgestellt [20]. Der Fokus dieser Arbeit liegt auf der Herleitung des zur Positronen-Emissions-Tomographie gehörendem Messoperators  $G$  und der Modellierung eines Minimierungsproblems, das die Schlechtgestellttheit des Minimierungsproblems 1.1 umgeht. Das Vorgehen ist wie folgt aufgeteilt:

Wir beginnen mit der Einführung maßtheoretischer und funktionalanalytischer Grundlagen und zeigen ein für unsere Anwendung wichtiges Korollar des Rieszschen Darstellungssatzes. Innerhalb der Grundlagen wird anschließend die direkte Methode vorgestellt als Leitfaden für den Existenzbeweis von Lösungen einer Funktion. Mit der Stochastischen Modellierung und der Definition von Poisson-Punkt-Prozessen erhalten wir die letzten theoretischen Werkzeuge, die der Beschreibung des Messvorgangs dienlich sind.

Nach einer kurzen Erläuterung der physikalischen Funktionsweise der Positronen-Emissions-Tomographie wenden wir uns in Kapitel 3 der Herleitung des Messoperators  $G$  zu. Dafür konstruieren wir zunächst den Maße transformierenden Vorwärtoperator und eine Funktion, die den statischen Messvorgang beschreibt. Für die Messfunktion orientieren wir uns an den von Wirth et al. in „Dynamic Cell Imaging in PET with Optimal Transport Regularization“ [3] vorgestellten Messoperator. Anschließend führen wir das Phasenraum-Modell von Alberti et al. ein und stellen das zugehörige Minimierungsproblem auf.

Mittels der Maximum-a-posteriori Schätzung entwickeln wir in Kapitel 4 zunächst eine erste Variante eines Rekonstruktionsfunktional, werden jedoch anschließend in einer kurzen Analyse sehen, dass aufgrund der Formulierung Probleme auftreten die der exakten Rekonstruktion hinderlich sind. Zur Umgehung der Probleme leiten wir auf Basis der ersten Variante und des gewichteten Maximums-Funktional ein angepasstes Rekonstruktionsfunktional her. In Kapitel 5 betrachten wir dieses im Hinblick auf die Existenz von Minimierern. Die in den Grundlagen vorgestellte Direkte Methode wird hierfür genutzt. Mit der Vorstellung eines numerischen Verfahrens für die Berechnung solcher Minimierer schließt die Masterthesis ab. Dafür stellen wir die nötigen Diskretisierungen und den Frank-Wolfe-Algorithmus vor, wobei ein Überblick über die Implementierung und numerischen Resultate den Schlusspunkt setzt.



## 2 Grundlagen

Wir beginnen mit der Einführung nützlicher Definitionen und Sätze, auf die im Verlauf der Arbeit zurückgegriffen werden. Dabei stehen weniger die Beweise im Vordergrund, als viel mehr der Zusammenhang zwischen Theorie und Anwendung. Mit den Konzepten der Maßtheorie und Funktionalanalysis werden Radon-Maße und ihre Verbindung zum Dualraum von stetigen Funktionen erläutert. Der Zusammenhang erlaubt uns die Betrachtung der direkten Methode als Leitfaden für den Beweis der Existenz von Lösungen eines Minimierungsfunktionals, das wir in späteren Kapiteln herleiten werden. Eine wichtige Rolle bei der Bestimmung des zum Positronen-Emissions-Tomographie-Scanners gehörenden Messoperators spielen die stochastischen Ereignisse. Die bei der PET eingesetzten radioaktiven Nuklide zerfallen jeweils gemäß einer Poisson-Verteilung und erhobene Messungen beruhen letztendlich auf diesem Zufall. Die Modellierung des Messprozesses muss dementsprechend ausgerichtet sein, wobei das Augenmerk auf den Poisson-Punkt-Prozessen liegt. Zuvor betrachten wir wie auf Lösungen eines inversen Problems zu schließen ist, wenn Daten und Lösungen stochastischer Natur sind. Hierfür verschafft der Maximum-a-posteriori Schätzer Abhilfe, über den Lösungen von Minimierungsproblemen, beziehungsweise Inversen Problemen gefunden werden können.

### 2.1 Konzepte der Maßtheorie und Funktionalanalysis

In der angewandten Mathematik werden für das Modellieren von sich bewegenden Teilchen häufig Maße benutzt, die je nach Anwendung andere Eigenschaften und Strukturen aufweisen. Auch bei der PET-Bildrekonstruktion werden endliche Radon-Maße betrachtet, deren Definition wir uns in diesem Abschnitt anfangs zuwenden. Um eine Struktur zu erhalten, auch um verschiedene Radon-Maße miteinander vergleichen zu können, wird anschließend als wichtige Charakterisierung die Totale-Variations-Norm erläutert. Über den Rieszschen Darstellungssatz erhalten wir als Folgerung den interessanten funktionalanalytischen Zusammenhang zwischen Radon-Maßen und dem Dualraum stetiger Funktionen.

Als Quelle dienen die Bücher von Jürgen Elstrodt [6] und Rudin [19]. Zweiteres beschäftigt sich mit komplexwertigen Maßen, deren Aussagen stets auch für reellwertige Maße gelten.

Sei im Folgenden ein Hausdorff-Raum  $X$  gegeben, versehen mit der Borel- $\sigma$ -Algebra  $\mathcal{B}(X)$ . Zunächst stellen wir Definitionen von verschiedenen Klassen von Maßen vor.

**Definition 2.1.** Sei  $(A_i)_{i \in \mathbb{N}} \subset \mathcal{B}(X)$  eine beliebige Folge von paarweise disjunkten Mengen und  $(X, \mathcal{B}(X))$  ein Messraum.

(i) Ein *signiertes Maß* ist eine Abbildung

$$\mu : \mathcal{B}(X) \longrightarrow [-\infty, \infty) \quad , \text{ oder } \quad \mu : \mathcal{B}(X) \longrightarrow (-\infty, \infty] \quad ,$$

für die  $\mu(\emptyset) = 0$  und die  $\sigma$ -Additivität gilt, d.h.

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i) \quad .$$

(ii) Ein (*positives*) *Maß* ist eine Abbildung

$$\mu : \mathcal{B}(X) \longrightarrow [0, \infty] \quad ,$$

mit der Eigenschaft der  $\sigma$ -Additivität und  $\mu(\emptyset) = 0$ .

(iii) Ein (signiertes) Maß  $\mu$  bezeichnen wir als *endliches (signiertes) Maß*, wenn gilt:

$$|\mu(X)| < \infty \quad .$$

Den Definitionen nach ist jedes positive Maß ein signiertes Maß. Außerdem folgt aus der  $\sigma$ -Additivität für endliche Maße, dass die Summe über  $\mu(A_i)$  absolut konvergiert. Grund hierfür ist die Invarianz der Reihenfolge der Mengen  $(A_i)_i$ . Im weiteren Verlauf betrachten wir positive Maße und bezeichnen diese einfachheitshalber als Maß, ohne den Zusatz positiv. Maße, die auf der Borelschen  $\sigma$ -Algebra definiert sind, werden auch *Borelmaße* genannt, jedoch herrscht in der Fachliteratur Uneinigkeit bezüglich der Definition von Borelmaßen. So fordern manche Autoren für Borelmaße zusätzlich die sogenannte *lokale Endlichkeit*. Diese und weitere Klassifikationen von Maßen erläutern folgende Definitionen.

**Definition 2.2.** Sei  $\mu$  ein Maß auf dem Messraum  $(X, \mathcal{B}(X))$ .

(i) Das Maß  $\mu$  heißt *lokal endlich*, wenn für jedes  $x \in X$  eine offene Menge  $U \subset X$  mit  $x \in U$  existiert, so dass  $\mu(U) < \infty$  gilt.

(ii) Mit  $\sigma$ -*endlich* wird ein Maß bezeichnet, für das eine abzählbare Menge  $(A_i)_{i \in \mathbb{N}} \subset \mathcal{B}(X)$  mit  $\mu(A_i) < \infty$  für alle  $i \in \mathbb{N}$  existiert, die  $X$  überdeckt.

(iii) Das Maß  $\mu$  heißt von *innen regulär*, wenn

$$\mu(A) = \sup\{\mu(K) \mid K \subset A, K \text{ kompakt}\} \quad \text{für alle } A \in \mathcal{B}(X)$$

und von *außen regulär*, wenn

$$\mu(A) = \inf \{ \mu(U) \mid A \subset U, U \text{ offen} \} \quad \text{für alle } A \in \mathcal{B}(X) .$$

Insgesamt bezeichnen wir ein Maß, das sowohl von innen als auch von außen regulär ist, als *reguläres Maß*.

Wir übernehmen die Definition nach der ein Borelmaß ein lokal endliches Maß ist, das auf der borelschen  $\sigma$ -Algebra agiert. Offensichtlich gilt, dass jedes endliche Maß auch  $\sigma$ -endlich ist.

Wie anfangs erwähnt fällt in den späteren Kapiteln den sogenannten Radon-Maßen eine besondere Rolle zu.

**Definition 2.3.** Ein lokal endliches und von innen reguläres Maß heißt *Radon-Maß*.

Dementsprechend ist ein Radon-Maß auf  $\mathcal{B}(X)$ , das zusätzlich zu der inneren Regularität auch von außen regulär ist, ein reguläres Borelmaß. Im Allgemeinen gilt, falls der Grundraum  $X$  ein topologischer, lokalkompakter Hausdorff-Raum mit abzählbarer Basis ist, versehen mit der zugehörigen Borel- $\sigma$ -Algebra, dass alle endlichen Borelmaße regulär und somit auch alle endlichen Radon-Maße regulär sind.

Bekannte Beispiele für Radon-Maße auf dem  $\mathbb{R}^n$  sind das Lebesgue-Maß und das Dirac-Maß, definiert durch

$$\delta_x(A) := \mathbb{1}_A(x) = \begin{cases} 1 & \text{falls } x \in A \\ 0 & \text{sonst} \end{cases} ,$$

für  $x \in \mathbb{R}^n$  und  $A \in \mathcal{B}(\mathbb{R}^n)$ . Ebenso sind auch Linearkombinationen von Dirac-Maßen wieder Radon-Maße. Genau diese dienen uns später für die Modellierung von Positionen verschiedener Teilchen innerhalb eines Teilgebietes  $[0, 1]^d$  des  $\mathbb{R}^d$ . Der Hyperwürfel  $[0, 1]^d$  beschreibt hierbei das durch einen PET-Tomographen abgedeckte Gebiet.

Daher wollen wir die Struktur der Menge aller (endlichen) Radon-Maße genauer verstehen und definieren diese für einen topologischen Hausdorff-Raum  $X$ , versehen mit einer lokal-kompakten Topologie, durch

$$M(X) := \{ \mu : \mathcal{B}(X) \longrightarrow [0, \infty] \mid \mu \text{ endliches Radon-Maß auf } X \} .$$

Dass  $M(X)$  ein Vektorraum ist, wird in [8, Prop. 7.16] bewiesen. Mit der sogenannten Totalen-Variation sind wir in der Lage, den Raum  $M(X)$  zusätzlich mit einer Norm zu versehen.

**Definition 2.4.** Sei  $\mu \in M(X)$  und die Abbildung  $|\mu|$  gegeben durch

$$\begin{aligned} |\mu| : \mathcal{B}(X) &\longrightarrow [0, \infty) \\ A &\longmapsto |\mu|(A) := \sup \left\{ \sum_i |\mu(A_i)| \mid (A_i)_i \subset \mathcal{B}(X) \text{ Partition von } A \right\} , \end{aligned}$$

wobei eine Partition von  $A$  eine Menge  $(A_i)_i$  von paarweise disjunkten Mengen  $A_i \in \mathcal{B}(X)$  ist, so dass  $A = \cup_i A_i$ . Der Abbildungsvorschrift entsprechend ist  $|\mu|$  ein positives, endliches Maß auf  $(X, \mathcal{B}(X))$  und heißt *Totales-Variations-Maß*. Mit

$$\|\mu\|_{TV} := |\mu|(X)$$

bezeichnen wir die *Totale-Variation* von  $\mu$  und allgemein für Maße aus  $M(X)$  mit

$$\begin{aligned} \|\cdot\|_{TV} : M(X) &\longrightarrow [0, \infty) \\ \mu &\longmapsto |\mu|(X) \end{aligned}$$

die *Totale-Variations-Norm*, abgekürzt auch *TV-Norm*.

Eine wichtige Eigenschaft, die es zu zeigen gilt, ist die Vollständigkeit von  $(M(X), \|\cdot\|_{TV})$ . Hierfür nutzen wir aus, dass die TV-Norm auf  $M(X)$  eine Metrik induziert und demnach es ausreicht zu zeigen, dass jede Cauchy-Folge in  $M(X)$  konvergiert. Zuvor wird jedoch der Satz von Radon-Nikodym benötigt. Dafür stellen wir zunächst die Lebesgue-Räume, das Bildmaß eines Maßes unter einer Funktion, ebenso wie einen zugehörigen Transformationssatz vor.

Für bezüglich eines Maßes  $\mu$  messbare Funktionen  $f : X \rightarrow \mathbb{R}$  sei der  $\mathbb{R}$ -Vektorraum

$$\mathcal{L}_\mu^p(X) := \{f : X \rightarrow \mathbb{R} \mid f \text{ } \mu\text{-messbar und } \|f\|_p < \infty\}$$

definiert, versehen mit der Halbnorm

$$\|f\|_p := \left( \int_X |f|^p d\mu \right)^{\frac{1}{p}}, \quad p \geq 1.$$

Dass  $(\mathcal{L}_\mu^p(X), \|\cdot\|_p)$  ein halbnormierter Raum und kein normierter Raum ist, liegt an der fehlenden Definitheit. So folgt aus  $\|f\| = 0$  nicht, dass es sich bei  $f$  um die Nullfunktion handelt. Definieren wir jedoch die Menge

$$\mathfrak{N} := \{f \in \mathcal{L}_\mu^p(X) \mid f = 0 \text{ } \mu\text{-fast überall}\}$$

und anschließend den Faktorraum

$$\mathcal{L}_\mu^p(X) / \mathfrak{N} =: L_\mu^p(X),$$

so erhalten wir den Lebesgue-Raum  $L_\mu^p(X)$ , auf dem nun  $\|\cdot\|_p$  eine Norm bildet.

**Definition 2.5.** Sei  $\mu$  ein Maß auf  $(X, \mathcal{B}(X))$ . Weiter sei  $(\tilde{X}, \mathcal{B}(\tilde{X}))$  ein weiterer Hausdorff-Raum und mit  $f : X \rightarrow \tilde{X}$  eine Borel-messbare Abbildung gegeben. Die Vorschrift

$$\begin{aligned} f_{\#}\mu : \mathcal{B}(\tilde{X}) &\longrightarrow [0, \infty] \\ A &\longmapsto f_{\#}\mu(A) := \mu(f^{-1}(A)) \end{aligned}$$

ist ein Maß auf  $\tilde{X}$  und heißt das *Bildmaß von  $\mu$  unter  $f$* .

Das Bildmaß eines Maßes unter einer Funktion wird in der Fachliteratur von manchen Autoren auch *Pushforward* genannt. Bei uns findet der Pushforward Anwendung bei der Modellierung von sich linear bewegenden Teilchen im Phasenraum. Eine wichtige Eigenschaft, die wir in diesem Kontext benötigen, ist die Transformationsformel für Bildmaße: Für eine  $f_{\#}\mu$ -messbare Abbildung  $g : \tilde{X} \rightarrow \mathbb{R}$  gilt:

$$\int_{\tilde{X}} g \, d(f_{\#}\mu) = \int_X g \circ f \, d\mu . \quad (2.1)$$

Ferner definieren wir für eine  $\mu$ -messbare Funktion  $h : X \rightarrow [0, \infty)$  die *Gewichtung von  $\mu$  mit  $h$*  durch

$$(h \cdot \mu)(A) := \int_A h \, d\mu , \quad A \in \mathcal{B}(X) . \quad (2.2)$$

Nun sind wir in der Lage, einen zentralen Satz der Maßtheorie vorzustellen. Er gibt Auskunft darüber, unter welchen Bedingungen ein Maß auch als Gewichtung eines anderen Maßes mit einer Funktion geschrieben werden kann.

**Satz 2.6.** (*Satz von Radon-Nikodym*)

Seien  $\mu$  ein  $\sigma$ -endliches und  $\lambda$  ein  $\sigma$ -endliches, signiertes Maß auf  $(X, \mathcal{B}(X))$ . Dann sind äquivalent:

(i)  $\lambda$  ist absolut stetig bezüglich  $\mu$ , d.h.

$$\mu(A) = 0 \implies \lambda(A) = 0 \quad \text{für alle } A \in \mathcal{B}(X) ,$$

notiert mit  $\lambda \ll \mu$ .

(ii) Es existiert eine nicht negative  $\mu$ -messbare Funktion  $f : X \rightarrow \mathbb{R}$ ,  $f \in L^1_{\mu}(X)$ , mit

$$\lambda = f \cdot \mu ,$$

also der Eigenschaft

$$\lambda(A) = \int_A f \, d\mu \quad \text{für alle } A \in \mathcal{B}(X).$$

Die Funktion  $f$  ist eindeutig durch  $\mu$  und  $\lambda$  bestimmt und heißt *Radon-Nikodym-Dichte von  $\lambda$  bezüglich  $\mu$* .

*Beweis.* Rudin beweist bei [19, Thm. 6.10] eine Version des Satzes von Radon-Nikodym für komplexe Maße.  $\square$

Wie bereits erwähnt ist ein endliches Radon-Maß  $\mu \in M(X)$   $\sigma$ -endlich. Für  $|\mu|$  als Totales-Variations-Maß gemäß Definition 2.4 betrachte eine  $|\mu|$ -Nullmenge  $A \in \mathcal{B}(X)$ . Es gilt

$$0 = |\mu|(A) = \sup \left\{ \sum_i |\mu(A_i)| \mid (A_i)_i \subset \mathcal{B}(X) \text{ Partition von } A \right\},$$

woraus sofort folgt, dass

$$|\mu(A_i)| = 0 \quad \text{für alle } A_i \in \mathcal{B}(X), (A_i)_i \subset \mathcal{B}(X) \text{ Partition von } A, \quad (2.3)$$

und demnach

$$\mu(A) = \mu\left(\bigcup_i A_i\right) = \sum_i \mu(A_i) = 0.$$

Dies beweist allgemein die absolute Stetigkeit von  $\mu$  bezüglich des zugehörigen Totalen-Variations-Maßes  $|\mu|$ . Nach dem Satz von Radon-Nikodym existiert folglich eine Funktion  $f \in L^1_{|\mu|}(X)$ , so dass  $\mu = f \cdot |\mu|$  gilt.

Definieren wir nun für  $r > 0$  die Menge

$$A_r := \{x \in X \mid |f(x)| < r\} \subset X$$

und mit  $(A_{r_i})_i$  eine Partition von  $A_r$ , so folgt die Abschätzung

$$\begin{aligned} \sum_i |\mu(A_{r_i})| &= \sum_i \left| \int_{A_{r_i}} f \, d|\mu| \right| \leq \sum_i \int_{A_{r_i}} |f| \, d|\mu| < \sum_i \int_{A_{r_i}} r \, d|\mu| \\ &= \sum_i r |\mu|(A_{r_i}) = r |\mu|\left(\bigcup_i A_{r_i}\right) \\ &= r |\mu|(A_r) \end{aligned}$$

und insgesamt gilt

$$\begin{aligned} |\mu|(A_r) &= \sup \left\{ \sum_i |\mu(A_{r_i})| \mid (A_{r_i})_i \subset \mathcal{B}(X) \text{ Partition von } A_r \right\} \\ &< r |\mu|(A_r). \end{aligned}$$

Für  $r = 1$  ergibt sich aus der obigen Ungleichung

$$|\mu|(A_1) = |\mu|\left(\{x \mid |f(x)| < 1\}\right) = 0.$$

Demnach ist  $A_1$  eine  $|\mu|$ -Nullmenge und für  $f$  gilt  $|f(x)| \geq 1$   $|\mu|$ -fast überall. Andererseits gilt für alle  $A \in \mathcal{B}(X)$  mit  $|\mu|(A) > 0$ :

$$\left| \frac{1}{|\mu|(A)} \int_A f \, d|\mu| \right| \stackrel{2.6}{=} \frac{|\mu(A)|}{|\mu|(A)} \leq 1.$$

Es folgt  $|f(x)| \leq 1$   $|\mu|$ -fast überall (siehe [19, Thm. 1.40]). Mit den Abschätzungen nach oben und unten für  $f$  erhalten wir insgesamt  $|f(x)| = 1$   $|\mu|$ -fast überall. Durch das undefinieren von  $f$  hin zu  $|f(x)| = 1$  für alle  $x \in X$ , haben wir eine Folgerung des Satzes von Radon-Nikodym hergeleitet, die auch als Polarzerlegung bekannt ist.

**Satz 2.7.** Sei  $\mu$  ein  $\sigma$ -endliches Maß, dann existiert eine Funktion  $f \in L^1_{|\mu|}(X)$  mit

$$|f(x)| = 1 \quad \text{für } x \in X \quad \text{und} \quad \mu = f \cdot |\mu| .$$

Die Polarzerlegung erlaubt es, für komplexe Maße deren Integration zu definieren, jedoch beschäftigen wir uns weiterhin mit reellwertigen Radon-Maßen und gehen nicht weiter auf die Theorie komplexwertiger Maße ein.

Eine weitere Anwendung des Satzes von Radon-Nikodym findet sich in dem folgenden Beweis des essentiellen Satzes über die Vollständigkeit des Raumes  $M(X)$ .

**Satz 2.8.** Sei  $X$  ein topologischer Hausdorff-Raum, versehen mit einer lokalkompakten Topologie, und  $M(X)$  der Raum aller endlichen Radon-Maße auf  $X$ . Dann ist  $(M(X), \|\cdot\|_{TV})$  ein vollständiger, normierter Vektorraum, also ein Banachraum.

*Beweis.* Wir zeigen nur die Vollständigkeit.

Sei  $(\mu_k)_k \subset M(X)$  eine Cauchy-Folge. Es gilt  $\sup_{k \in \mathbb{N}} \|\mu_k\|_{TV} < \infty$  und die Folge ist demnach beschränkt in  $M(X)$ . Definiere

$$\mu := \sum_{k=1}^{\infty} \frac{1}{2^k} |\mu_k| .$$

Dann ist  $\mu$  nach Konstruktion ein endliches Radon-Maß auf  $X$  und Element von  $M(X)$ . Sei  $A \in \mathcal{B}(X)$  eine beliebige Nullmenge bezüglich  $\mu$ . Da alle  $|\mu_k|$  nicht-negativ sind, ist  $A$  auch eine  $\mu_k$ -Nullmenge für alle  $k \in \mathbb{N}$ . Es folgt die absolute Stetigkeit aller  $\mu_k$  bezüglich  $\mu$  und nach Satz 2.6 existieren für alle  $k \in \mathbb{N}$  Funktionen  $(f_k)_k \subset L^1_{\mu}(X)$  mit  $\mu_k = f_k \cdot \mu$ . Insbesondere folgt damit, dass

$$\begin{aligned} \|\mu_k - \mu_l\|_{TV} &= \|f_k \cdot \mu - f_l \cdot \mu\|_{TV} \\ &= \sup \left\{ \sum_i |(f_k \cdot \mu - f_l \cdot \mu)(A_i)| \mid (A_i)_i \text{ Partition von } X \right\} \\ &= \sup \left\{ \sum_i (|f_k - f_l| \cdot \mu)(A_i) \mid (A_i)_i \text{ Partition von } X \right\} \\ &= \sup \left\{ \sum_i \int_{A_i} |f_k - f_l| d\mu \mid (A_i)_i \text{ Partition von } X \right\} \\ &= \int_X |f_k - f_l| d\mu = \|f_k - f_l\|_1 . \end{aligned}$$

Der Gleichung nach, und mit der Charakterisierung von  $(\mu_k)_k$  als Cauchy-Folge, ist auch  $(f_k)_k$  eine Cauchy-Folge in  $L^1_{\mu}(X)$ . Der Satz von Riesz-Fischer besagt, dass  $L^1_{\mu}(X)$  die Struktur eines Banachraums aufweist und aufgrund dessen Vollständigkeit konvergiert  $(f_k)_k$  gegen eine Grenzfunktion  $f_{\infty} \in L^1_{\mu}(X)$ . Wir definieren

$$\mu_{\infty} := f_{\infty} \cdot \mu = \sum_{k=1}^{\infty} \frac{1}{2^k} (f_{\infty} \cdot |\mu_k|) \in M(X)$$

und es folgt mit den gleichen Rechenschritten wie zuvor:

$$\begin{aligned}\|\mu_\infty\|_{TV} &= \|f_\infty\|_1 < \infty, \\ \|\mu_k - \mu_\infty\|_{TV} &= \|f_k - f_\infty\|_1 \longrightarrow 0.\end{aligned}$$

Folglich konvergiert  $(\mu_k)_k$  in  $M(X)$  gegen  $\mu_\infty$ , was die Vollständigkeit von  $M(X)$  beweist.  $\square$

Wir haben mit Satz 2.8 die für uns später sehr wichtige Eigenschaft der Vollständigkeit festgehalten. Im nächsten Schritt wollen wir  $M(X)$  mit Hilfe des Rieszschen Darstellungssatzes mit einem Dualraum identifizieren, den wir im Folgenden einführen. Zuvor benötigen wir weitere Definitionen, unter anderem die Einführung des Raumes aller beschränkten, bzw. des Raumes aller im Unendlichen verschwindenden Funktionen.

**Definition 2.9.** Sei  $X$  ein topologischer Raum und  $f : X \longrightarrow \mathbb{R}^d$  eine (stetige) Funktion.

- (i) Die Funktion  $f$  *verschwindet im Unendlichen*, falls für alle  $\epsilon > 0$  eine kompakte Menge  $K \subset X$  existiert, sodass für alle  $x \in X \setminus K$  die Funktionswerte von  $f$  betragsweise kleiner als  $\epsilon$  sind, das heißt  $|f(x)| < \epsilon$  für alle  $x \in X \setminus K$ .
- (ii) Mit

$$\text{supp}(f) := \overline{\{x \in X \mid f(x) \neq 0\}}$$

wird der *Träger* von  $f$  bezeichnet. Der Träger ist auch für nicht-stetige Funktionen definiert.

Diesen Definitionen nach fassen wir die Menge aller Funktionen mit den besagten Eigenschaften wie folgt zusammen.

**Definition 2.10.** Sei  $X$  hausdorffsch mit lokalkompakter Topologie und  $C(X)$  der Raum aller stetigen Funktionen  $f : X \longrightarrow \mathbb{R}^d$ . Wir notieren drei Unterräume von  $C(X)$  durch:

- (i) Mit  $C_b(X)$  bezeichnen wir den *Raum aller beschränkten, stetigen Funktionen*.
- (ii) Den *Raum aller im Unendlichen verschwindenden Funktionen* notieren wir mit  $C_0(X)$ .
- (iii)  $C_c(X)$  sei der *Raum aller stetigen Funktionen mit kompaktem Träger*.

Durch die Abbildungsvorschrift

$$\begin{aligned}\|\cdot\|_\infty : C(X) &\longrightarrow \mathbb{R} \\ f &\longmapsto \|f\|_\infty := \sup_{x \in X} |f(x)|\end{aligned}$$

können wir die Teilräume  $C_b(X)$ ,  $C_c(X)$  und  $C_0(X)$  mit einer Norm versehen. Es gilt, dass  $(C_0(X), \|\cdot\|_\infty)$  ein Banachraum ist. Für den Beweis sei auf [19, Thm. 3.17] verwiesen. Falls, wie in späterer Anwendung der Fall,  $X$  eine kompakte Teilmenge des  $\mathbb{R}^n$  ist, gilt die Gleichheit zwischen allen Räumen:  $C_b(X) = C_c(X) = C_0(X)$ .

Wir stellen nun den Darstellungssatz von Riesz vor. Seine wesentliche Aussage besagt, dass jedes lineare und beschränkte Funktional auf  $C_0(X)$  eine Darstellung durch Integration bezüglich eines eindeutigen, endlichen Radon-Maßes besitzt. Jedoch gibt es verschiedene Versionen des Satzes: Je nachdem, welche Grundräume betrachtet werden, besitzen die darstellenden Maße andere Eigenschaften. Zum Beispiel muss  $C_c(X)$  für das zugrundeliegende Radon-Maß nicht endlich sein, für  $C(X)$ , mit  $X$  zusätzlich  $\sigma$ -kompakt, besitzen die Radon-Maße kompakte Träger.

Wir nehmen die Aussage über  $C_0(X)$  als Grundlage, wobei in der späteren Anwendung ein Funktionenraum über Kompakta herangezogen wird.

**Satz 2.11.** (*Darstellungssatz von Riesz*)

*Sei  $X$  ein lokalkompakter Hausdorff-Raum, versehen mit der zugehörigen Borel- $\sigma$ -Algebra  $\mathcal{B}(X)$  und  $(C_0(X))'$  der topologische Dualraum von  $C_0(X)$ .*

*Für alle  $\phi \in (C_0(X))'$  (lineares und stetiges Funktional,  $\phi : C_0(X) \rightarrow \mathbb{R}$ ) existiert ein eindeutiges Maß  $\mu \in M(X)$  mit*

$$\phi(f) = \int_X f \, d\mu \quad \text{für alle } f \in C_0(X) .$$

*Für die Operatornorm von  $\phi$  gilt zudem:*

$$\|\phi\|_{op} = \|\mu\|_{TV} .$$

Die Operatornorm ist eine auf  $(C_0(X))'$  agierende Norm, die den Dualraum zu einem Banachraum macht. Sie ist definiert durch

$$\|\phi\|_{op} := \sup \left\{ |\phi(f)| \mid \|f\|_\infty = 1, f \in C_0(X) \right\} .$$

*Beweis.* Siehe [19, Thm. 6.19] . □

Der Darstellungssatz von Riesz wird auch als Darstellungssatz von Riesz-Markow-Kakutani betitelt, benannt nach den Autoren, die entscheidend zu seiner Entdeckung beigetragen haben. Wir stellen als nächsten Schritt den Zusammenhang zwischen  $M(X)$  und  $(C_0(X))'$  her.

**Satz 2.12.** *Sei  $T : M(X) \rightarrow (C_0(X))'$  eine Abbildung, charakterisiert durch*

$$\begin{aligned} \mu &\longmapsto T_\mu \quad \text{mit der Auswertung} \\ T_\mu(f) &:= \int_X f \, d\mu \quad \text{für } f \in C_0(X) . \end{aligned}$$

Dann ist  $T$  ein isometrischer Isomorphismus und für die TV-Norm eines Maßes  $\mu \in M(X)$  gilt:

$$\|\mu\|_{TV} = \sup \left\{ \left| \int_X f \, d\mu \right| \mid \|f\|_\infty = 1, f \in C_0(X) \right\}.$$

*Bemerkung 2.13.* Eigentlich folgt aus dem Satz von Riesz für beliebige topologische Räume  $X$  lediglich, dass der Raum aller regulären Maße isometrisch isomorph zu dem Dualraum von  $C_0(X)$  ist. Doch durch die Wahl eines lokalkompakten Hausdorff-Raum mit zugehöriger Borel- $\sigma$ -Algebra sind die regulären Maße auf  $X$  genau die Radon-Maße. Daher behält der obige Satz seine Gültigkeit.

*Beweis.* (Skizzenhaft)

Aus dem Rieszschen Darstellungssatz folgt, dass  $T$  bijektiv und isometrisch ist. Weiter gilt für alle  $f \in C_0(X)$

$$|T_\mu(f)| \leq \left| \int_X f \, d\mu \right| \leq \int_X |f| \, d|\mu| \leq \|f\|_\infty \|\mu\|_{TV} < \infty$$

und  $T$  ist dementsprechend wohldefiniert. Aus der Isometrie und erneut aus Satz 2.11 folgt:

$$\begin{aligned} \|\mu\|_{TV} &= \|T_\mu\|_{op} = \sup \left\{ |T_\mu(f)| \mid \|f\|_\infty = 1, f \in C_0(X) \right\} \\ &= \sup \left\{ \left| \int_X f \, d\mu \right| \mid \|f\|_\infty = 1, f \in C_0(X) \right\}. \end{aligned}$$

□

Zusammengefasst wissen wir nun, dass  $M(X)$  ein Banachraum ist, der mit dem Dualraum von  $C_0(X)$  identifiziert werden kann. Demnach ist es möglich abschließend die schwach- $*$ -Konvergenz für Folgen von Maßen in  $M(X)$  zu definieren.

**Definition 2.14.** Eine Folge von Maßen  $(\mu_n)_n \subset M(X)$  ist *schwach- $*$ -konvergent* gegen  $\mu \in M(X)$ , falls

$$\int_X f \, d\mu_n \longrightarrow \int_X f \, d\mu \quad \text{für alle } f \in C_0(X).$$

Wir notieren diese Konvergenz durch  $\mu_n \xrightarrow{w^*} \mu$ .

## 2.2 Direkte Methode

In diesem Kapitel wird ein bewährtes und einfaches Vorgehen vorgestellt, um die Existenz von Minimierern eines Funktionals zu beweisen. Dabei geben wir notwendige Eigenschaften an das Funktional und an den Raum wider, aus dem die Argumente des Funktionals stammen, die zu einer Existenzaussage führen.

Das Vorgehen im Detail ist Folgendes: Zunächst betrachten wir den Satz von Banach-Alaoglu, leiten die schwach-\* -Unterhalbstetigkeit der TV-Norm her und beschreiben anschließend die direkte Methode und ihre Bedingungen. Die Inhalte orientieren sich dabei an Kapitel drei des Buches von Irene Fonesca und Giovanni Leoni [11].

**Definition 2.15.** Eine Menge  $A \subset X$  eines topologischen Raumes  $X$  heißt präkompakt, falls ihr Abschluss,  $\overline{A} \subset X$ , kompakt in  $X$  ist.

**Satz 2.16.** (Banach-Alaoglu) Sei  $Z$  ein Banachraum und  $Y = Z'$  sein topologischer Dualraum. Falls eine Teilmenge  $B \subset Y$  beschränkt bezüglich der Norm in  $Y$  ist, dann ist  $B$  bezüglich der schwach-\* -Topologie präkompakt.

*Beweis.* Siehe [21, S.335 f.] □

Der Satz von Banach-Alaoglu verdeutlicht, dass ein enger Zusammenhang zwischen bezüglich der Operatornorm beschränkten Mengen in einem Dualraum und der schwach-\* -Konvergenz von (Teil-)Folgen innerhalb der Menge besteht. Denn angewandt bedeutet die Aussage von Banach-Alaoglu, dass für bezüglich der Operatornorm beschränkte Folgen eine schwach-\* -konvergente Teilfolge existiert.

Der nächste Satz ebnet den Weg für die schwach-\* -Unterhalbstetigkeit der TV-Norm.

**Satz 2.17.** Sei  $X$  ein Banachraum,  $\|\cdot\|_X$  die zugehörige Norm, und  $(\phi_n)_n \subset X'$  eine Folge mit  $\phi_n \xrightarrow{w^*} \phi$  in  $X'$ . Es gelten:

(i) Die schwach-\* -konvergente Folge  $(\phi_n)_n$  ist beschränkt bezüglich der Operatornorm auf dem Raum  $X'$ .

(ii) Die Operatornorm auf  $X'$  ist schwach-\* -unterhalbstetig, das heißt

$$\|\phi\|_{op} \leq \liminf_{n \rightarrow \infty} \|\phi_n\|_{op} .$$

*Beweis.* Die Folge  $(\phi_n)_n$  ist beschränkt bezüglich der schwach-\* -Topologie auf  $X'$ . Demnach gilt die punktweise Beschränktheit

$$\sup \left\{ |\phi_n(x)| \mid n \in \mathbb{N} \right\} < \infty \quad \text{für alle } x \in X ,$$

und es folgt aus dem Prinzip der gleichmäßigen Beschränktheit sofort

$$\sup \left\{ \|\phi_n\|_{op} \mid n \in \mathbb{N} \right\} < \infty .$$

Für die zweite Aussage betrachte die folgende Abschätzung, die für alle  $x \in X$  gilt:

$$|\phi(x)| = \lim_{n \rightarrow \infty} |\phi_n(x)| = \liminf_{n \rightarrow \infty} |\phi_n(x)| \leq \liminf_{n \rightarrow \infty} \|\phi_n\|_{op} \|x\|_X$$

womit nach Umstellen und für  $\|x\|_X \neq 0$  folgt

$$\frac{|\phi(x)|}{\|x\|_X} \leq \liminf_{n \rightarrow \infty} \|\phi_n\|_{op} .$$

Für die Operatornorm von  $\phi$  gilt somit

$$\begin{aligned} \|\phi\|_{op} &= \sup \left\{ |\phi(x)| \mid \|x\|_X = 1, x \in X \right\} \\ &= \sup \left\{ \frac{|\phi(x)|}{\|x\|_X} \mid \|x\|_X \neq 0, x \in X \right\} \\ &\leq \liminf_{n \rightarrow \infty} \|\phi_n\|_{op} . \end{aligned}$$

□

Seien im Folgenden  $X$  ein lokalkompakter Banachraum,  $C_0(X)$  und  $M(X)$ , versehen mit der TV-Norm, als Dualraum gegeben.

**Satz 2.18.** *Die TV-Norm ist schwach- $*$ -unterhalbstetig auf  $M(X)$ .*

*Beweis.* Aus Satz 2.11, Satz 2.12 und Satz 2.17 folgt die Aussage, denn die Operatornorm auf  $(C_0(X))'$  ist gleich der TV-Norm. □

Die theoretischen Grundlagen für die direkte Methode sind nun bekannt. Mit einer Funktion  $J : X \rightarrow \mathbb{R}$  wollen wir diese vorstellen, wobei das Ziel darin besteht zu zeigen, dass  $J$  einen (eindeutigen) Minimierer  $x \in X$  besitzt. Dafür reichen vier Schritte aus:

Schritt 1: Betrachte eine Minimalfolge in  $X$ , d.h. eine Folge  $(x_n)_n \subset X$  für die gilt:

$$\lim_{n \rightarrow \infty} J(x_n) = \inf_{x \in X} J(x) .$$

Schritt 2: Zeige, dass eine Teilfolge  $(x_{n_k})_k \subset X$  von  $(x_n)_n$  existiert, die bezüglich einer, möglicherweise schwächeren, Topologie  $\tau$  gegen ein Element  $x_0 \in X$  konvergiert.

Schritt 3: Beweise die Unterhalbstetigkeit der Funktion  $J$  bezüglich der Topologie  $\tau$ .

Schritt 4: Aus den Schritten eins bis drei folgt für  $J$  bereits die Existenz eines Minimierers, für die Eindeutigkeit sollte das Funktional  $J$  noch strikt konvex sein.

Die ersten beiden Schritte stellen im Allgemeinen keine großen Herausforderungen dar: Für die Existenz einer Minimalfolge reicht die Wohldefiniertheit von  $J$  und die Beschränktheit von unten in Form einer Konstante  $C \in \mathbb{R}$  aus, so dass  $J(x) \geq C$  für alle  $x \in X$  gilt. Für den zweiten Schritt ist es hilfreich zu zeigen, dass  $J$  eine koerzive Funktion ist. Dadurch wird garantiert, dass die Minimalfolge beschränkt ist. Zusammen mit dem Satz von Banach-Alaoglu kann dann auf die Existenz einer schwach-\*-konvergenten Teilfolge geschlossen werden und der Grenzwert dieser Teilfolge ist ein Kandidat für den Minimierer von  $J$ . Schwierigkeiten bereitet jedoch der dritte Schritt, denn im Allgemeinen ist  $J$  nicht unterhalbstetig, auch nicht bezüglich einer schwächeren Topologie  $\tau$ .

Um dieses Problem zu lösen wird auf eine relaxierte Funktion  $J'$  zurückgegriffen, die mit  $J$  in dem Sinne eng verwandt ist, als dass sie folgende Eigenschaften besitzt:

- (i)  $J'$  ist unterhalbstetig bezüglich  $\tau$  (in unserer Anwendung bezüglich der schwach-\*-Topologie).
- (ii)  $J'$  ist koerziv.
- (iii) Es gilt die Gleichheit:

$$\min_{x \in X} J'(x) = \inf_{x \in X} J(x) . \quad (2.4)$$

Mit diesen Eigenschaften erfüllt  $J'$  die Voraussetzungen der direkten Methode. Demnach existiert ein Minimum und wir können statt dem Infimum von  $J'$  das Minimum schreiben.

Die Funktion  $J'$  besitzt also alle notwendigen Bedingungen und das Minimum ist nach Gleichung (2.4) identisch mit dem Infimum von  $J$ . Für  $J'$  können wir insgesamt die direkte Methode anwenden und das Vorgehen wird im folgenden Satz derart festgehalten, dass es der späteren Anwendung entspricht.

Zuvor gilt es aber zu beachten, dass durch den Beweis der Existenz von Minimierern für  $J'$  nicht die Existenz von Minimierern von  $J$  folgt. Im Allgemeinen gilt sogar, dass selbst im Falle der Existenz für beide Funktionen die Minimierer nicht gleich sind. Mit der Relaxierung  $J'$  haben wir lediglich eine Formulierung vorliegen, deren Minimierer auch bestimmbar sind und denen von  $J$  (falls sie existieren) „ähneln“.

**Satz 2.19.** (*Direkte Methode*)

Sei  $X$  ein lokalkompakter, normierter Hausdorff-Raum und  $M(X)$  der Raum der Radonmaße auf  $X$ . Sei  $(\mu_n)_n \subset M(X)$  eine Folge und weiter  $J : M(X) \rightarrow \mathbb{R}$  eine wohldefinierte Funktion mit folgenden Eigenschaften:

- (i) Es existiert eine Konstante  $C \in \mathbb{R}$  mit  $J(\mu) \geq C$  für alle  $\mu \in M(X)$ .

(ii)  $J$  ist koerziv, d.h.

$$\|\mu_n\|_{TV} \rightarrow \infty \implies J(\mu_n) \rightarrow \infty .$$

(iii)  $J$  ist schwach- $*$ -unterhalbstetig, also

$$\mu_n \xrightarrow{w^*} \mu \in M(X) \implies \liminf_{n \rightarrow \infty} J(\mu_n) \geq J(\mu) .$$

Dann existiert ein Minimierer von  $J$ . Ist  $J$  zudem noch strikt konvex, so ist der Minimierer eindeutig.

*Beweis.* Durch die Beschränktheit von unten können wir auf die Existenz einer Minimalfolge  $(\mu_n)_n \subset M(X)$  schließen, die gegen die untere Schranke konvergiert:

$$\lim_{n \rightarrow \infty} J(\mu_n) = \inf_{\mu \in M(X)} J(\mu) \geq C .$$

Dementsprechend gilt für die Minimalfolge  $J(\mu_n) \nrightarrow \infty$  und aus der Koerzivität von  $J$  folgt  $\|\mu_n\|_{TV} \nrightarrow \infty$ . Somit existiert eine Konstante  $M \geq 0$  mit

$$\|\mu_n\|_{TV} < M < \infty \quad \text{für alle } n \in \mathbb{N} .$$

Die Minimalfolge bildet daher eine bezüglich der TV-Norm beschränkte Menge in  $M(X)$ . Nach dem Satz von Banach-Alaoglu (Satz 2.16) ist  $(\mu_n)_n \subset M(X)$  als Menge bezüglich der schwach- $*$ -Topologie präkompakt und es existiert eine Teilfolge  $(\mu_{n_k}) \subset M(X)$  die schwach- $*$  gegen einen Grenzwert  $\mu_0 \in M(X)$  konvergiert. Nun folgt aus der schwach- $*$ -Unterhalbstetigkeit von  $J$ , dass

$$\liminf_{k \rightarrow \infty} J(\mu_{n_k}) \geq J(\mu_0) ,$$

und insgesamt

$$J(\mu_0) \geq \inf_{\mu \in M(X)} J(\mu) = \lim_{n \rightarrow \infty} J(\mu_n) = \lim_{k \rightarrow \infty} J(\mu_{n_k}) = \liminf_{k \rightarrow \infty} J(\mu_{n_k}) \geq J(\mu_0) .$$

Aus dieser Ungleichungskette erschließt sich, dass  $J$  sein Minimum bei  $\mu_0$  annimmt:

$$J(\mu_0) = \inf_{\mu \in M(X)} J(\mu) = \min_{\mu \in M(X)} J(\mu) .$$

Es bleibt die Eindeutigkeit des Minimierers zu zeigen. Hierfür nehmen wir an, dass zwei Minimierer  $\mu_1, \mu_2$  von  $J$  mit  $\mu_1 \neq \mu_2$  existieren und konstruieren einen neuen Kandidaten für den Minimierer. Für  $\alpha \in (0, 1)$  und mit der strikten Konvexität folgt:

$$J(\alpha\mu_1 + (1 - \alpha)\mu_2) < \alpha J(\mu_1) + (1 - \alpha)J(\mu_2) = J(\mu_1) = J(\mu_2) = \min_{\mu \in M(X)} J(\mu) .$$

Demnach ist  $J(\alpha\mu_1 + (1 - \alpha)\mu_2)$  echt kleiner als das Minimum von  $J$ , was im Widerspruch steht. Der Minimierer von  $J$  ist daher eindeutig.  $\square$

Wir haben in diesem Kapitel einen für uns sehr nützlichen Satz hergeleitet. Für ein gegebenes Rekonstruktionsfunktional reicht es demnach aus, die nötigen Voraussetzungen nachzurechnen, und der Beweis der Existenz eines eindeutigen Minimierers ist vollzogen. Leider wird sich herausstellen, dass die zuerst hergeleitete Minimierungsfunktion nicht die Eigenschaften der Konvexität und Unterhalbstetigkeit erfüllt. Wir greifen daher auf eine relaxierte Funktion, wie in diesem Kapitel beschrieben, zurück.

## 2.3 Stochastische Modellierung

In der Medizinischen Bildgebung mittels der Positronen-Emissions-Tomographie unterliegen die Messungen stochastischen Schwankungen. So wird der Zerfall eines Positrons durch eine entsprechende Verteilung bestimmt. Des Weiteren können auch stets Messfehler auftreten, die ebenfalls einer zugehörigen Verteilung gehorchen. In der Modellierung durch ein Inverses Problem stellen diese Zufälle neue Anforderungen an die Rekonstruktion. Um wünschenswerte Ergebnisse zu erhalten, wird Vorwissen benötigt, das durch geschickte Formulierung in die Rekonstruktion eingeht. Das Vorgehen dieses Kapitels orientiert sich an die Vorlesung von Benedikt Wirth, gehalten während des Sommersemesters 2019, [2] und an die Vorlesung von Martin Burger, Wintersemester 2016/2017, [4]. Des Weiteren dient eine Arbeit von Antonin Chambolle et. al. [1] dem Verständnis.

Wir betrachten ein Inverses Problem  $K(x) = y$ , das heißt für einen gegebenen Operator  $K : \mathbb{R}^n \rightarrow \mathbb{R}^m$  und rechte Seite  $y \in \mathbb{R}^m$  ist  $x \in \mathbb{R}^n$  gesucht, sodass  $K(x) = y$ . Die exakte Lösung  $x^*$  ist die Realisierung einer Zufallsvariable  $X$ . In  $X$  steckt das a-priori-Wissen um die Lösungen, indem dieses in die Modellierung der Verteilungsfunktion von  $X$  eingeht. Weiter modellieren wir die Daten  $y$  als Realisierung einer Zufallsvariable  $Y$ , für die  $\mathbb{E}(Y) = y = K(x^*)$  gilt. Wie wahrscheinlich ist es nun, die Daten  $y$  zu messen, wenn  $x^*$  als Lösung bekannterweise vorliegt?

Dadurch, dass  $X$  und  $Y$  Zufallsvariablen sind und  $x, y$  jeweilige Realisierungen, können wir diese Wahrscheinlichkeit durch die bedingte Wahrscheinlichkeit von  $y$  bezüglich  $x$  ausdrücken:

$$\mathbb{P}[Y = y | X = x^*] = \frac{\mathbb{P}[Y = y, X = x^*]}{\mathbb{P}[X = x^*]}, \text{ wobei } \mathbb{P}[X = x^*] \neq 0.$$

*Bemerkung 2.20.* Falls  $X$  und  $Y$  voneinander unabhängig sind, so gilt

$$\mathbb{P}[Y = y, X = x] = \mathbb{P}[Y = y]\mathbb{P}[X = x] \quad \text{und demnach } \mathbb{P}[Y = y | X = x] = \mathbb{P}[Y = y].$$

In der Anwendung von realen Inversen Problemen ist die Unabhängigkeit jedoch fast nie gegeben.

Eine weitere Möglichkeit, die bedingte Wahrscheinlichkeit auszudrücken, bietet der Satz von Bayes.

**Satz 2.21.** (*Satz von Bayes*)

Seien  $X$  und  $Y$  Zufallsvariablen und  $\mathbb{P}[X = x] > 0$ . Dann gilt

$$\mathbb{P}[Y = y|X = x] = \frac{\mathbb{P}[X = x|Y = y]\mathbb{P}[Y = y]}{\mathbb{P}[X = x]} .$$

*Beweis.* Siehe [2]. □

Der Satz von Bayes stellt einen Zusammenhang zwischen der bedingten Wahrscheinlichkeit von  $y$ , gegeben  $x$  und  $x$  gegeben  $y$  her. Demnach erlaubt er uns die Fragestellung umzudrehen: Wie wahrscheinlich ist das Vorliegen der Lösung  $x$ , wenn wir die Daten  $y$  gemessen haben?

Mit dieser Frage befinden wir uns mitten im Inversen Problem. Die Mathematik hat hierfür zwei Ansätze entwickelt, zum einen den Maximum-Likelihood-Schätzer, zum anderen den Maximum-a-posteriori Schätzer.

**Definition 2.22.** Betrachte das inverse Problem zu gegebenen Messungen/Daten  $y$  die Lösung  $x$  zu bestimmen.

(i) Der *Maximum-Likelihood (ML) Schätzer* für  $x$  ist definiert durch

$$x^* = \operatorname{argmax}_X \mathbb{P}[Y = y|X = x] = \operatorname{argmin}_X -\log \mathbb{P}[Y = y|X = x] .$$

(ii) Der *Maximum-a-posteriori (MAP) Schätzer* für  $x$  ist definiert durch

$$x^* = \operatorname{argmax}_X \mathbb{P}[X = x|Y = y] = \operatorname{argmin}_X -\log \mathbb{P}[Y = y|X = x] - \log \mathbb{P}[X = x] .$$

*Bemerkung 2.23.* Die Notation  $\operatorname{argmin}_X$  bedeutet, dass wir nach einer Realisierung  $x$  der Zufallsvariable  $X$  suchen, die den Term minimiert. Ferner folgt die Gleichheit zwischen  $\operatorname{argmax}_X$  und  $\operatorname{argmin}_X$  für den ML- und MAP-Schätzer aus dem Satz von Bayes und der Anwendung des Logarithmus. Beispielhaft gilt für den MAP-Schätzer gilt ( $\mathbb{P}[Y = y] \neq 0$ ):

$$\begin{aligned} \operatorname{argmax}_X \mathbb{P}[X = x|Y = y] &= \operatorname{argmin}_X -\log \mathbb{P}[X = x|Y = y] \\ &= \operatorname{argmin}_X -\log \frac{\mathbb{P}[Y = y|X = x]\mathbb{P}[X = x]}{\mathbb{P}[Y = y]} \\ &\stackrel{2.21}{=} \operatorname{argmin}_X -\log \mathbb{P}[Y = y|X = x] - \log \mathbb{P}[X = x] + \log \mathbb{P}[Y = y] . \end{aligned}$$

Da wir über  $X$  minimieren, wird der Term  $\log \mathbb{P}[Y = y]$  weggelassen und wir erhalten die Gleichheit in Definition 2.22 (ii).

Es stellt sich heraus, dass der MAP-Schätzer als Rekonstruktionsmethode besser geeignet ist, im Vergleich zu dem ML-Schätzer. Um dies zu sehen, verdeutlichen wir uns kurz, was genau der Unterschied zwischen den beiden Vorgehensweisen ist.

Der ML-Schätzer wählt das  $x$  aus, das mit der höchsten Wahrscheinlichkeit die Daten  $y$  erzeugt hat. Dies führt dazu, dass der ML-Schätzer die bestapproximierendste Lösung annimmt. Auf Grund der Schlechtgestellttheit des vorliegenden Inversen Problems wird jedoch eine weitere, nicht genauer spezifizierte, Regularisierung benötigt.

Hingegen bestimmt der MAP-Schätzer die Lösung  $x$ , die am wahrscheinlichsten unter der beobachteten Messung  $y$  vorliegt. Zwar benötigen wir hier Vorwissen über die Verteilung von  $x$ , da der Term  $\mathbb{P}[X = x]$  eine Rolle spielt, aber dieses kann durch die Implementierung einer Regularisierung Teil der Rekonstruktion sein. Demnach beinhaltet der MAP-Schätzer bereits eine für die Regularisierung benötigte Struktur.

In der Anwendung modellieren wir die a-priori-Verteilung von  $X$  mit Hilfe des sogenannten *Gibbs-Priors*, wobei  $R(\cdot)$  dem eigentlichen Regularisierungsterm entspricht:

$$\mathbb{P}[X = x] = \exp(-\beta R(x)) , \text{ mit dem Regularisierungsparameter } \beta > 0 .$$

*Bemerkung 2.24.* Bei der PET-Bildrekonstruktion versuchen wir die Verteilung von zuvor verabreichten radioaktiv markierten Zellen im Körper  $X$  eines Untersuchungsobjektes zu bestimmen. Demnach modellieren wir die Zellen durch ein Maß  $\lambda^*$ , definiert durch eine Summe von Dirac-Maßen, wobei jedes Dirac-Maß den Ort einer Zelle beschreibt. Als Summe über Dirac-Maße ist  $\lambda^* \in M(X)$  selbst ein dünnbesetztes Maß. Mit diesem a-priori-Wissen bietet sich eine Regularisierung durch die TV-Norm an, denn diese erzwingt Dünnbesetztheit (siehe [2]). Wir wählen also  $R(\cdot) = \|\cdot\|_{TV}$  und erhalten insgesamt für gemessene Daten  $y$  (als Realisierung von  $Y$ ) den MAP-Schätzer:

$$\lambda^* \in \underset{\lambda \in M(X)}{\operatorname{argmin}} -\log \mathbb{P}[Y = y|\lambda] + \beta \|\lambda\|_{TV} .$$

Im Verlauf dieser Arbeit wird auf ein von Alberti et al. [9] vorgeschlagenes Minimierungsproblem zurückgegriffen. Ohne an dieser Stelle näher auf die Formulierung dessen einzugehen zeigt sich, dass die Suche nach einer Lösung durch den MAP-Schätzer ein geeignetes Verfahren beschreibt, um die Schlechtgestellttheit des Problems zu umgehen. Im Kontext zu vorherigem Kapitel wird dabei das zu minimierende Funktional  $J(\lambda)$ , mit der Struktur

$$J(\lambda) = -\log \mathbb{P}[Y = y|\lambda] + \beta \|\lambda\|_{TV} ,$$

auf seine Eigenschaften bezüglich Beschränktheit, Koerzivität, Konvexität und schwach\*-Unterhalbstetigkeit untersucht. Erfüllt  $J$  die Voraussetzungen, so können wir durch Satz 2.19 auf die Existenz einer  $J$  minimierenden Lösung schließen.

## 2.4 Poisson-Punkt-Prozesse

Während wir mit dem MAP-Schätzer einen Ansatz für die Lösung von inversen Problemen/Minimierungsproblemen besitzen, benötigen wir für die Modellierung des Messprozesses ein Hilfsmittel, das die Zufälligkeit des Zerfalls von Positronen als Kerneigenschaft der Messungen beschreibt. Dabei spielt der Ort des Zerfalls eine ebenso wichtige Rolle wie die hinter dem Emittier-Ereignis liegende Wahrscheinlichkeitsverteilung. Es wird sich herausstellen, dass Poisson-Punkt-Prozesse besonders gut geeignet sind, da die Zerfallswahrscheinlichkeit eines Nuklids in der Physik durch eine Poisson-Verteilung beschrieben wird. Als Einleitung in das Thema dient das folgende Beispiel, in dem eine Möglichkeit vorgestellt wird, wie eine Formulierung der Messfunktion aussehen kann und welche Eigenschaften diese aufweisen sollte.

Angenommen es befinden sich  $N$  durch radioaktive Isotope markierte Zellen in einem kompakten Gebiet  $X \subset \mathbb{R}^d$ . Jede der  $N$  Zellen emittiert während eines Messzeitraums  $[0, T]$ ,  $T > 0$ , zufällig insgesamt  $e_i$  Positronen. Es geschehen somit  $e_1 + \dots + e_N$  Ereignisse zu zufälligen Zeitpunkten. Um die Messungen zu beschreiben, gilt es festzuhalten, in welchem Teilraum von  $X$  wie viele Ereignisse geschehen sind. Einfacher beschrieben wird eine Funktion benötigt, die die Anzahl an Ereignissen zählt. Bildhaft können wir uns vorstellen, dass jedes Emittieren eines Positrons durch einen Punkt in  $X$  dargestellt wird. Über den Messzeitraum  $[0, T]$  hinweg füllt sich  $X$  folglich mit Punkten auf und die gewünschte Messfunktion soll alle Punkte in einem Bereich von  $X$  zählen, die zum Zeitpunkt  $T$  insgesamt entstanden sind. Mathematischer lässt sich dies durch eine Zufallsvariable  $Z$  beschreiben, deren Realisierung  $Z(\tilde{\omega})$  die Zerfallspunkte in  $X$  festlegt (durch das Abbilden auf die abzählbaren Teilmengen von  $X$ ), ferner mit einer Funktion  $\eta$ , deren Argumente  $\tilde{\omega}$  und Teilräume  $B$  von  $X$  sind,

$$\eta : (\tilde{\Omega}, \mathcal{F}, \mathbb{P}) \times \{B | B \subset X\} \longrightarrow \mathbb{N}_0$$

$$\tilde{\omega} \longmapsto \eta(\tilde{\omega}, B) := |Z(\tilde{\omega}) \cap B| .$$

Hierbei ist  $(\tilde{\Omega}, \mathcal{F}, \mathbb{P})$  der zum Zerfall der Isotope gehörende Wahrscheinlichkeitsraum, dessen genaue Struktur unbekannt bleibt. Es gilt  $|Z(\tilde{\omega})| = e_1(\tilde{\omega}) + \dots + e_N(\tilde{\omega})$ , ferner  $\eta(\tilde{\omega}, X) = e_1(\tilde{\omega}) + \dots + e_N(\tilde{\omega})$ , und für ein festes  $\tilde{\omega}$  ist  $\eta(\tilde{\omega}, \cdot)$  offensichtlich ein Maß, das auch *Zählmaß bezüglich  $Z$*  genannt wird. Genau dieses Maß modelliert den Messprozess in unserem Beispiel, entspricht also der gesuchten Funktion. Eine letzte wünschenswerte Eigenschaft für  $\eta$  wäre eine Erweiterung im folgenden Sinne: Anstatt nach dem Zeitpunkt  $T$  die Anzahl an Punkten zu messen, sollte  $\eta$  über die Zeit indiziert werden und die über einen Zeitraum  $[t_k, t_j] \subset [0, T]$  neu hinzugekommen Punkte zählen,  $0 \leq t_k < t_j \leq T$ . Dadurch werden auch die zufälligen Zerfallszeitpunkte erfasst und gemessen. Eine intuitive Lösung besteht darin,  $Z$  ebenfalls über unterschiedliche Zeitintervalle zu indizieren und durch die Definition von  $\eta_t(\tilde{\omega}, B) := |Z_t(\tilde{\omega}) \cap B|$  wird eine Unterteilung in mehrere Messzeiträume möglich. Hierbei steht  $t$  für einen Möglichen Messzeitraum  $[t_k, t_j]$ .

Das hergeleitete  $\eta$  entspricht in dem Beispiel der Definition eines Punkt-Prozesses, die wir nun in allgemeinerer Form vorstellen. Nach einer Charakterisierung über Intensitätsmaße und dem Erläutern der Campbell Formel, werden abschließend die für uns wichtigen Poisson-Punkt-Prozesse eingeführt. Dabei dient [10] als Grundlage für dieses Kapitel, doch zuvor

noch eine entscheidende Anmerkung zu obigem Beispiel:

Auch wenn  $\eta$  hier als Messfunktion vorgestellt wurde, so entspricht  $\eta$  in der Anwendung nicht der Messung im eigentlichen Sinne. Viel mehr beschreibt  $\eta$  lediglich die mathematische Erfassung von Zerfällen in  $X$  und ist Ziel einer Rekonstruktion. Dies liegt an der Funktionsweise eines PET-Scanners: Der Zerfall eines Nuklids wird indirekt an den Detektoren gemessen, als Punkt am Rand von  $X$  und nicht in  $X$ . Die Zählung dieser Punkte entspricht daher der wahren Messung, und mit deren Hilfe gilt es die Realisierung von  $Z$  zu finden, die letztendlich Auslöser für die von  $\eta$  gezählten Punkte ist.

Sei  $X$  ein topologischer Raum versehen mit der Borel- $\sigma$ -Algebra  $\mathcal{B}(X)$ . Weiter sei der Raum aller endlichen Maße auf  $X$  gegeben durch

$$N(X) := \{ \mu : \mathcal{B}(X) \rightarrow \mathbb{R} \mid \mu \text{ ist Maß auf } X \text{ und es gilt: } \mu(A) \in \mathbb{N}_0 \text{ für alle } A \in \mathcal{B}(X) \} .$$

Beispielhaft sind das triviale Null-Maß und Dirac-Maße Elemente von  $N(X)$ . Mit  $\mathcal{N}$  als Menge aller Teilmengen von  $N(X)$

$$\mathcal{N}(X) = \left\{ \{ \mu \in N(X) \mid \mu(A) = k \} \mid A \in \mathcal{B}(X), k \in \mathbb{N}_0 \right\},$$

ist eine  $\sigma$ -Algebra auf  $N(X)$  gegeben. Der Konstruktion nach ist  $\mathcal{N}$  die kleinste  $\sigma$ -Algebra auf  $N(X)$ , die alle Abbildungen  $\mu \mapsto \mu(A)$  messbar macht. Für einen Wahrscheinlichkeitsraum  $(\tilde{\Omega}, \mathcal{F}, \mathbb{P})$  kann nun die Definition eines Punktprozesses eingeführt werden.

**Definition 2.25.** Ein *Punktprozess* auf einem Messraum  $(X, \mathcal{B}(X))$  ist eine messbare Abbildung

$$\begin{aligned} \eta : \tilde{\Omega} &\longrightarrow (N(X), \mathcal{N}) \\ \tilde{\omega} &\longmapsto \eta(\tilde{\omega}, \cdot) := \eta(\tilde{\omega})(\cdot) , \end{aligned}$$

das bedeutet,  $\eta$  ist eine  $N(X)$ -wertige Zufallsvariable.

Definieren wir für  $A \in \mathcal{B}(X)$  die Funktion

$$\begin{aligned} \eta(A) : \tilde{\Omega} &\longrightarrow \mathbb{N}_0 \\ \tilde{\omega} &\longmapsto \eta(\tilde{\omega}, A) , \end{aligned}$$

so gilt für  $k \in \mathbb{N}_0$ , dass

$$\{ \eta(A) = k \} = \{ \tilde{\omega} \in \tilde{\Omega} \mid \eta(\tilde{\omega}, A) = k \} \in \mathcal{F} .$$

Dann ist  $\eta(A)$  nach Konstruktion eine Zufallsvariable mit Werten in  $\mathbb{N}_0$  und wir bezeichnen die Anzahl an Punkten von  $\eta$  in der Menge  $A$  als  $\eta(A)$ . Dies entspricht der Herleitung des Zählmaßes bezüglich  $Z$  in unserem Beispiel. Abhängig davon, welches  $\tilde{\omega} \in \tilde{\Omega}$  vorliegt, werden die Zerfallsereignisse (als Punkte in  $X$ ) gemessen.

Eine Charakteristik von Punktprozessen wird durch deren Intensitätsmaß definiert. Es beschreibt die erwartete Anzahl an Punkten eines Punktprozesses, die in beliebigen messbaren Mengen liegen.

**Definition 2.26.** Das *Intensitätsmaß* eines Punktprozesses auf  $(X, \mathcal{B}(X))$  ist ein Maß  $\lambda$ , das für  $A \in \mathcal{B}(X)$  definiert ist durch

$$\lambda(A) = \mathbb{E} [\eta(A)] .$$

Dabei sorgen die Eigenschaften des Erwartungswertes für die Charakterisierung von  $\lambda$  als Maß.

Einen wichtigen Zusammenhang zwischen Punktprozessen und deren Intensitätsmaßen stellt Campbells Formel her.

**Satz 2.27.** (*Campbells Formel*)

Sei  $\eta$  ein Punktprozess auf  $(X, \mathcal{B}(X))$  mit Intensitätsmaß  $\lambda$  und  $f : X \rightarrow \mathbb{R}$  eine messbare Funktion. Dann ist  $\int_X f d\eta(x)$  eine Zufallsvariable und es gilt

$$\mathbb{E} \left[ \int_X f d\eta(x) \right] = \int_X f d\lambda(x) ,$$

falls  $f$  nicht-negativ oder  $\int_X |f| d\lambda(x) < \infty$  ist.

*Beweis.* Siehe [10, Prop. 2.7]. □

Beschreiben wir den späteren Messoperator mittels eines (Poisson-) Punkt-Prozesses, so wird das hergeleitete Rekonstruktionsfunktional nach einem Punktprozess suchen, der sich den gemessenen Daten in dem Sinne nähert, als dass sein Erwartungswert den Messungen entspricht. An dieser Stelle hilft uns Satz 2.27, denn die Formel von Campbell charakterisiert die Berechnung des Erwartungswertes von Punktprozessen.

Zuletzt führen wir die Poisson-Punkt-Prozesse ein.

**Definition 2.28.** Sei  $\lambda$  ein  $\sigma$ -endliches Maß auf  $X$ . Ein *Poisson-Punkt-Prozess* mit Intensitätsmaß  $\lambda$  ist ein Punktprozess  $\eta$  auf  $(X, \mathcal{B}(X))$ , für den gilt:

(i) Für alle  $A \in \mathcal{B}(X)$  gilt

$$\eta_\lambda(A) := \eta(\cdot)(A) \stackrel{d}{=} \text{Poi}(\lambda(A)) .$$

(ii) Für alle  $A_1, \dots, A_m \in \mathcal{B}(X)$  mit  $A_i \cap A_j = \emptyset$  für  $i \neq j$  sind  $\eta_\lambda(A_1), \dots, \eta_\lambda(A_m)$  unabhängige Zufallsvariablen.

Nehmen wir nun für die Anzahl an radioaktiven Zerfällen innerhalb eines Zeitfensters und Gebietes die Poisson-Verteilung als Grundlage, so ist der zugehörige Punkt-Prozess gleich einem Poisson-Punkt-Prozess. Dadurch bietet sich für unsere Modellierung der Messungen nachher der Poisson-Punkt-Prozess bestens an.

## 3 Der Messoperator auf dem Phasenraum

Beginnend mit der Funktionsweise und Anwendungsbeispielen der Positronen-Emissions-Tomographie, werden in diesem Kapitel die wichtigsten Bausteine des mathematisch modellierten Messvorgangs hergeleitet. Zu den Bausteinen gehören zum einen der Maß transformierende Vorwärtsoperator, der Phasenraum mitsamt zugehörigem Modell und die Modellierung der zufälligen Mess-Ereignisse durch einen Poisson-Punkt-Prozess. Abschließend mündet die Kombination aller im dynamischen Messoperator. Dabei werden viele Konzepte erst allgemein eingeführt und im Laufe der Arbeit detaillierter erläutert. Als dem Verständnis hilfreiche Lektüre sind die Paper von Lee et al.[14] und Kristian Bredies [15] hervorzuheben, ebenso wie die Masterthesis von Marco Mauritz [16].

### 3.1 Funktionsweise der Positronen-Emissions-Tomographie

Die Positronen-Emissions-Tomographie (PET) ist ein medizinisches Bildgebungsverfahren, das unter anderem Anwendung in der Onkologie und der Neurologie findet. Zu Beginn des Verfahrens erhält das Untersuchungsobjekt ein Pharmazeutikum, das mit kurzlebigen radioaktiven Nukliden versehen ist. Diese emittieren Positronen, auch  $\beta^+$ -Strahlung genannt, die im Körper in Wechselwirkung mit nahen Elektronen geraten. Sich treffende Positronen und Elektronen (Antiteilchen des Positrons) löschen sich gegenseitig aus und es entstehen zwei Photonen, die sich vom Punkt der Wechselwirkung aus mit Lichtgeschwindigkeit in entgegengesetzte Richtung fortbewegen. Dieser Vorgang heißt Annihilation. Mittels Detektoren erfasst ein PET-Tomograph im nächsten Schritt die beiden Photonen. Über einen vorher festgelegten Zeitraum werden somit indirekt die zufälligen Zerfälle der Nuklide gemessen. Im letzten Schritt werden die Daten durch einen Algorithmus ausgewertet und die Verteilung der Nuklide im Körper wird rekonstruiert.

Die Onkologie nutzt zum Beispiel die Eigenschaft von Tumoren aus viel Energie zum wachsen zu benötigen. Um die genaue Position eines Tumors zu lokalisieren, wird dem Untersuchungsobjekt ein Präparat aus radioaktiv markierter Glucose verabreicht. Der hohe Energiebedarf von Tumoren sorgt dafür, dass die Krebszellen die Glucose vermehrt aufnehmen. Dadurch entsteht eine Ansammlung von radioaktiven Isotopen im Bereich der Wucherung, deren indirekte Photonen-Strahlung (entstanden durch die Annihilation) sich mit einem PET-Scanner messen lässt und den Tumor schlussendlich lokalisierbar macht.

Ein weiteres Anwendungsbeispiel der PET ist die Untersuchung von Gefäßen/Adern von Menschen und Tieren. Hierfür werden dem Untersuchungsobjekt erneut radioaktiv markier-

te Zellen verabreicht, diesmal jedoch injiziert. Demnach bewegen sich die Zellen mit dem Blutfluss, oder bleiben an verengten Stellen hängen. Somit gilt es hier, neben der Position, auch die Geschwindigkeit der Zellen zu ermitteln. Nachdem über einen festgelegten Zeitraum gemessen wurde, gibt es verschiedene Ansätze aus diesen Messungen die Position und Geschwindigkeit einer jeden Zelle zu bestimmen.

Hier setzen wir an und untersuchen im Folgenden einen möglichen Ansatz für die simultane Rekonstruktion von Geschwindigkeiten und Positionen mehrerer Zellen. Dabei werden wir ein Model von Alberti et al. [9] mit der von Wirth et al. in [3] vorgestellten Methode kombinieren.

### 3.2 Der Vorwärtsoperator

Der erste Schritt bei der Herleitung eines geeigneten Rekonstruktionsverfahrens besteht in dem Finden einer Antwort zu folgender Frage: Welche Messungen erwarten wir für eine bekannte Verteilung der radioaktiven Substanz im Körper?

Demnach gilt es eine Funktion zu konstruieren, deren Argument ein Maß ist, operierend auf dem Gebiet des Tomographen und die als Output die zugehörige Messung liefert. Der Herleitung dieser Messfunktion gelten die kommenden Unterkapitel.

Mit  $\Omega \subset \mathbb{R}^d$ ,  $d \in \mathbb{N}$ , als kompaktes und konvexes Teilgebiet beschreiben wir die räumlichen Ausmaße des PET-Scanners. Das Untersuchungsobjekt befindet sich im Inneren von  $\Omega$ , wohingegen die Detektoren sich entlang des Randes  $\partial\Omega$  verteilen. Die Verteilung der radioaktiv markierten Teilchen im Körper modellieren wir durch ein Radonmaß  $\mu$ . Da sich in unserer Anwendung die Zellen über die Zeit hinweg bewegen, indizieren wir  $\mu$  via  $\mu_t$ , das nun die Verteilung der Nuklide zum Zeitpunkt  $t \in [0, T]$  festhält. Für  $T > 0$  ist hierbei  $[0, T]$  das Zeitfenster, während dessen gemessen wird. Somit ist das Maß insgesamt von der Form

$$\mu = dt \times \mu_t \in M([0, T] \times \Omega) , \tag{3.1}$$

wobei  $dt$  als Lebesgue-Maß die Zeit modelliert. Angemerkt sei, dass in der Anwendung selbstverständlich  $d = 3$  gilt. Die Betrachtung im  $d$ -Dimensionalen ist theoretischer Natur, erlaubt aber die Funktionstauglichkeit des Modells numerisch in zwei Dimensionen zu überprüfen.

Wie in 3.1 erklärt, wird  $\mu$  nur indirekt gemessen. Für die Modellierung der Messungen benötigen wir daher eine Funktion, die  $\mu_t$ , als Maß auf  $\Omega$ , zu einem Maß auf  $\partial\Omega \times \partial\Omega$  transformiert. Um dies zu verstehen, betrachten wir hier den Messvorgang genauer: Zu der Emittation eines Positrons gehören zwei Photonen und dementsprechend wird die Energie des Zerfalls an beide Photonen abgegeben. Diese transportieren die Energie weiter, bis sie jeweils auf einen Detektor stoßen, der diese Energie misst. Die Detektoren befinden sich am Rand von  $\Omega$ , daher findet die Messung auf  $\partial\Omega \times \partial\Omega$  statt. Für dieses Kapitel reicht es aus, den exakten Abdeckungsbereich eines Detektors allgemein zu halten.

Die Struktur der gesuchten Funktion ist somit gegeben und wir betiteln sie als *Vorwärts-*

operator:

$$F : M(\Omega) \longrightarrow M(\partial\Omega \times \partial\Omega)$$

$$\lambda \longmapsto F(\lambda) .$$

Als nächsten Schritt gilt es, auch das Lebesgue-Maß, als zentralen Bestandteil von  $\mu$ , zu transformieren. Hierfür nutzen wir die Identität, denn die Zeitpunkte/räume der Messungen bleiben gleich, unabhängig von der Transformation der Maße  $\mu_t$  hin zu  $F(\mu_t)$ . Wir erhalten für die Transformation von  $\mu$  die Abbildungsvorschrift

$$M([0, T] \times \Omega) \longrightarrow M([0, T] \times \partial\Omega \times \partial\Omega)$$

$$\mu = dt \times \mu_t \longmapsto dt \times F(\mu_t) .$$

Im Folgenden leiten wir die Abbildungsvorschrift von  $F$  her. Nach physikalischen Überlegungen ergeben sich drei verschiedene Wege, wie das bei der Annihilation freigesetzte Photonenpaar gemessen werden kann:

1. (Nichtmessung) Eines der (oder beide) Photonen wird nicht von den Detektoren erfasst.
2. (Streuung) Mindestens eines der beiden Photonen wird gestreut und daraus ergibt sich keine entgegengesetzte Flugrichtung. Beide Photonen werden jedoch nahezu zeitgleich von zwei Detektoren erfasst.
3. (Erfassung) Das Photonenpaar fliegt ohne Streuung perfekt in entgegengesetzte Richtung und wird von den Detektoren erfasst.

Dem nachkommend besteht der Vorwärtsoperator  $F$  aus drei Teilen

$$F = F^0 + F^S + F^E : M(\Omega) \longrightarrow M(\partial\Omega \times \partial\Omega) ,$$

die wir einzeln herleiten.

$F^0$ : Mit  $F^0$  beschreiben wir die Zerfälle, die nicht zu einer Messung führen. Mathematisch schickt  $F^0$  ein Maß  $\mu \in M(\Omega)$  entsprechend auf das triviale Nullmaß  $0 \in M(\Omega)$ ,

$$F^0 : M(\Omega) \longrightarrow M(\partial\Omega \times \partial\Omega)$$

$$\mu \longmapsto 0 .$$

Der Nulloperator spielt nur theoretisch eine Rolle und die Erwähnung dient ausschließlich der Vollständigkeit halber. Im Verlauf der Arbeit wird nicht weiter auf  $F^0$  eingegangen, wir behalten jedoch im Hinterkopf, dass nicht jeder Zerfall eines Nuklids gemessen werden muss. Ein möglicher Grund hierfür liegt an der Anordnung der Detektoren, da zwischen benachbarten Detektoren Zwischenräume bestehen.

$F^S$ : Für das Modellieren des Streuungsoperators nehmen wir an, dass die Streuung eines oder beider Photonen die Flugrichtung zufällig verändert. Dieser Annahme nach ist die

Wahrscheinlichkeit, dass ein Photonenpaar die Positionen  $(x, y) \in \partial\Omega \times \partial\Omega$  erreicht, homogen gleichverteilt über  $\partial\Omega \times \partial\Omega$  (Sie ist unabhängig von  $(x, y)$ ). Der Operator  $F^S$  verteilt  $\mu$  gleichmäßig um auf  $\partial\Omega \times \partial\Omega$ . Des Weiteren erhalten wir ein Maß auf  $\partial\Omega \times \partial\Omega$ , das als Hausdorff-Maß identifiziert wird und die Funktionsvorschrift von  $F^S$  ist wie folgt gegeben:

$$F^S : M(\Omega) \longrightarrow M(\partial\Omega \times \partial\Omega)$$

$$\mu \longmapsto \mu(\Omega) \frac{(H^{d-1} \times H^{d-1})_{(\partial\Omega \times \partial\Omega)}}{(H_{\partial\Omega}^{d-1}(\partial\Omega))^2},$$

wobei  $H^{d-1}$  das  $d - 1$ -dimensionale Hausdorff-Maß notiert. Beispielfhaft decken zwei Detektoren die Flächen  $A, B \subset \partial\Omega$  ab. Wir erhalten als Messung für das Detektorenpaar  $A \times B$  die Auswertung von  $F^S(\mu)(A \times B)$ :

$$F^S(\mu)(A \times B) = \underbrace{\mu(\Omega)}_{\text{Gesamtmasse von } \mu} \cdot \frac{(H^{d-1} \times H^{d-1})_{(\partial\Omega \times \partial\Omega)}(A \times B)}{\underbrace{(H_{\partial\Omega}^{d-1}(\partial\Omega))^2}_{\text{Anteil der von } A \times B \text{ abgedeckten Fläche an } (\partial\Omega \times \partial\Omega)}}.$$

Wenn eine weitere Annahme besagt, dass die Detektoren gleich groß sind, sprich  $H^{d-1}(A) = H^{d-1}(B)$  im vorherigen Beispiel gilt, so folgt, dass der Streuungsoperator tatsächlich für jedes Detektorenpaar die gleiche Intensität/Messung ermittelt.

$F^E$ : Die Modellierung des Operators, der den Erfassungsprozess ohne Streuung beschreibt, beruht auf der hinter dem Prozess stehenden Physik. Der Prozess wird eingeleitet durch das Lösen eines Positrons aus dem Radionuklid. Dieses legt einen (meistens sehr kurzen) Weg zurück bis es auf ein Elektron stößt, wobei der Weg im Allgemeinen vom umgebenden Gewebe abhängt. Mathematisch beschreiben wir die Wahrscheinlichkeit, dass ein an der Stelle  $x \in \Omega$  emittiertes Positron bei  $y \in \Omega$  auf ein Elektron trifft mit einer stetigen Funktion  $G : \Omega \times \Omega \longrightarrow [0, \infty)$ . Für ein festes  $x \in \Omega$  ist  $G_x : \Omega \longrightarrow [0, \infty)$  eine Wahrscheinlichkeitsdichte. Somit gibt  $G_x(A)$  die Wahrscheinlichkeit an, dass das bei  $x$  emittierte Positron auf ein Elektron im Bereich  $A \subset \Omega$  trifft. Für unsere Anwendung ignorieren wir das umliegende Gewebe und betrachten einen räumlich homogenen Gauß-Filter als Grundlage für  $G$ ,

$$G_x(y) := G(y, x) = \frac{1}{(\sqrt{2\pi}\epsilon)^2} \cdot \exp\left(-\frac{\|x - y\|_2^2}{2\epsilon^2}\right) \quad \text{für } \epsilon > 0 \text{ fest.}$$

Der erste Schritt des Erfassungsoperators transformiert gemäß Gaußscher Glockenkurve die Intensität eines Positrons, emittiert bei  $x$ , auf die Umgebung von  $x$ , mit Peak an der Stelle  $x$  selbst,

$$F_1^E : M(\Omega) \longrightarrow M(\Omega)$$

$$\mu \longmapsto \left( A \longmapsto \int_A \int_{\Omega} G_x(y) d\mu(x) dy \right).$$

Der Annihilation nach fliegen zwei Photonen in entgegengesetzte Richtung und beide befinden sich auf einer Linie. Die Wahrscheinlichkeitsverteilung der Ausrichtung dieser Linie ist

dabei homogen, sprich die Flugrichtung der Photonen ist gleichwahrscheinlich zufällig. Wir wollen im nächsten Schritt das Maß  $F_1^E(\mu)$  verknüpfen mit einem weiteren Maß, das diese Linie mitsamt ihren zufälligen Eigenschaften beschreibt. Hierfür betrachten wir mit  $G(1, d)$  die Graßmann-Mannigfaltigkeiten von ein-dimensionalen Unterräumen des  $\mathbb{R}^d$ . Auf  $G(1, d)$  existiert gemäß [18, Thm. 13.2.11, Thm. 13.2.12] ein eindeutiges, rotationsinvariantes Wahrscheinlichkeitsmaß  $v_d$ , Haarsches-Maß genannt. Somit erfüllt  $v_d$  mit diesen Eigenschaften die nötigen Voraussetzungen an die Beschreibung einer Linie und wir erhalten durch

$$\begin{aligned} F_2^E : M(\Omega) &\longrightarrow M(\Omega \times G(1, d)) \\ \mu &\longmapsto \mu \times v_d \end{aligned}$$

den zweiten Teil des Erfassungsoperators.

Der letzte Schritt hin zur Messung der Photonen ist das Auftreffen beider auf die Detektoren, beziehungsweise der Kontakt mit dem Rand von  $\Omega$ . Es gilt die Positionen dieses Kontaktes herzuleiten. Mit den Informationen über die Flugrichtung  $v \in G(1, d)$  und dem Annihilationsursprung  $x \in \Omega$  können wir die lineare Funktion  $f(z) = x + z\tilde{v}$  definieren, wobei  $z \in \mathbb{R}$  und  $\tilde{v} \in v$  ein beliebiges Element ist, das als Steigung dient. Die beiden Schnittpunkte von  $f$  mit  $\partial\Omega$  bezeichnen wir mit  $s_1$  und  $s_2$ ,

$$\{s_1, s_2\} = \{f(z) \in \mathbb{R}^d \mid z \in \mathbb{R}\} \cap \partial\Omega .$$

Dass genau zwei Schnittpunkte existieren, folgt aus der Kompaktheit und Konvexität von  $\Omega$ , und aus der Linearität von  $f$ . Demnach sind  $s_1$  und  $s_2$  die Positionen, an denen die Photonen detektiert werden. Durch

$$\begin{aligned} P : \Omega \times G(1, d) &\longrightarrow \partial\Omega \times \partial\Omega \\ (x, v) &\longmapsto (s_1, s_2) \end{aligned}$$

wird eine Abbildung definiert, die nach zuvor beschriebenem Vorgehen das Auftreffen der Photonen am Rand ermittelt. Die Abbildung  $P$  ermöglicht es uns, das Maß  $F_2^E(\mu)$  hin zu einem auf  $\partial\Omega \times \partial\Omega$  agierenden Maß zu transformieren. Mit Hilfe des Bildmaßes von  $F_2^E(\mu)$  unter  $P$ , siehe Definition 2.5, ist  $F^E$  gegeben, durch

$$\begin{aligned} F_3^E : M(\Omega \times G(1, d)) &\longrightarrow M(\partial\Omega \times \partial\Omega) \\ \mu &\longmapsto P\#\mu . \end{aligned}$$

Fügen wir nun alle Komponenten zusammen, so bildet

$$\begin{aligned} F^E : M(\Omega) &\longrightarrow M(\partial\Omega \times \partial\Omega) \\ \mu &\longmapsto F_3^E F_2^E F_1^E(\mu) \end{aligned}$$

den Operator, der die Detektion eines Zerfalls ohne Streuung der Photonen modelliert.

Zuletzt gilt es, die Ereignisse der Nicht-Messung, Streuung und perfekten Erfassung mit Wahrscheinlichkeiten zu gewichten. Ein Zerfall des Radionuklids führt schließlich nur zu einem der drei Ereignisse. Mit  $p_0, p_S, p_E : \Omega \longrightarrow [0, 1]$  seien drei Zufallsvariablen definiert, die

zu den entsprechenden Ereignissen gehören. Insgesamt gilt  $p_0 + p_S + p_E = 1$  und die Zufallsvariablen stehen in starker Abhängigkeit zum Ort des Zerfalls. Da diese Abhängigkeit jedoch schwer zu erfassen ist, nehmen wir raumzeitlich konstante Werte für die Wahrscheinlichkeiten an, verweisen aber auf die möglichen Optimierungen, die sich hier mit a-priori-Wissen ergeben (zum Beispiel werden im Unterarm annihilerte Photonen zu einer geringeren Wahrscheinlichkeit gestreut als im Oberschenkel).

Als Auswertung des Vorwärtsoperators  $F$  für ein Maß  $\mu \in M(\Omega)$  erhalten wir folglich:

$$F(\mu) = p_0 F^0(\mu) + p_S F^S(\mu) + p_E F^E(\mu) = p_S F^S(\mu) + p_E F^E(\mu) \in M(\partial\Omega \times \partial\Omega)$$

und insgesamt ist der Vorwärtsoperator gegeben durch:

$$\begin{aligned} F : M(\Omega) &\longrightarrow M(\partial\Omega \times \partial\Omega) \\ \lambda &\longmapsto p_S F^S(\lambda) + p_E F^E(\lambda) . \end{aligned} \tag{3.2}$$

*Bemerkung 3.1.* Auf eine weitere wichtige Eigenschaft von  $F$  sei bereits hier verwiesen:  $F$  ist linear. Dies liegt daran, dass  $F$  eine Verknüpfung von linearen Funktionen  $F^S$  und  $F_{1,2,3}^E$  ist. Die Linearität jener ist auf die Linearität der Integration bezüglich eines Maßes ( $F_1^E$ ), weiter auf die Linearität des Push-Forwards ( $F_3^E$ ) zurückzuführen und andernfalls trivialerweise gegeben.

### 3.3 Dynamische Super-Resolution

Ein intuitiver Rekonstruktionsansatz der Positionen und Geschwindigkeiten von radioaktiv markierten Zellen im Körper ist der Folgende:

Mit festgelegten Zeitabständen werden mehrere Messungen durchgeführt. Zu jedem Messzeitpunkt wird die Position der Zellen rekonstruiert, um anschließend den von den Zellen zurückgelegten Pfad zu ermitteln. Dafür wird zwischen den verschiedenen Scans interpoliert und mittels den bekannten Zeitabstände zwischen zwei Aufnahmen kann auf die Geschwindigkeit geschlossen werden. Dadurch, dass zunächst jede Messung einzeln ausgewertet, und erst im Nachhinein die Geschwindigkeit ermittelt wird, heißt dieses Vorgehen auch *statische Rekonstruktion*.

Auch wenn diese Methode Erfolg verspricht, birgt sie Nachteile, wie Alberti et al. in ihrem Paper [9] einwenden. Vor allem bemängeln Alberti et al. den Nicht-Nutzen aller Daten. Dies bedarf einer kurzen Erklärung: Bei der statischen Rekonstruktion werden für die Ermittlung der Positionen der Zellen zu einem Zeitpunkt die vor und nach dem Messzeitpunkt erfassten Daten nicht genutzt. Das bedeutet, dass die vorherigen Positionen der Zellen, ebenso wie die Nachherigen, nicht in der Positions-Rekonstruktion eines Zeitpunktes berücksichtigt werden. Und das, obwohl das Ausnutzen aller Daten genauere Bestimmungen ermöglichen kann.

Weiter heben Alberti et al. den Fall hervor, dass zu einem Messzeitpunkt die Zellen für eine Rekonstruktion zu Nahe beieinander liegen. Dadurch ist eine fehlerfreie (unter einer

Fehlertoleranz) Bestimmung der Zellen nicht möglich und Daten dieses Scans sind nicht auszuwerten. Eine Betrachtung zeitlich benachbarter Messungen würde jedoch die Möglichkeit bieten, (fast) exakt zu rekonstruieren, trotz zu nahe liegender Zellen zu einem Zeitpunkt. Hinsichtlich dieser Mängel schlagen die Autoren eine alternative Rekonstruktions-Idee vor. Alle zu jedem Zeitpunkt erhobenen Daten werden im sogenannten Phasenraum betrachtet. Die Informationen über Ort und Geschwindigkeit einer Zelle treten dort gepaart als Tupel auf, wodurch bei der Rekonstruktion über den Phasenraum, die Positionen und Geschwindigkeiten simultan ermittelt werden. Dabei wird von einer linearen Dynamik der Zellen während des gesamten Messprozesses ausgegangen. Bevor wir uns dem Raum-Geschwindigkeits-Modell und seiner Herleitung widmen, seien noch zwei Anmerkungen zu beachten.

Zunächst sticht die starke Einschränkung an die Bewegung der Zellen heraus. Eine lineare Dynamik ist sehr selten gegeben und in der Praxis auch nicht der Fall. Adern und Nervenbahnen, selbst Tumorzellen, nehmen keine lineare Bewegung an. Dennoch kann durch geeignete Modellierung der Messungen zumindest auf kurzfristige lineare Bewegungen geschlossen werden, indem der Messzeitraum kurz genug gewählt wird. Die Methode ist folglich anwendbar, auch wenn für die Rekonstruktion von nicht-linearen Verläufen Umformungen und Erweiterungen nötig sind. Eine Möglichkeit dennoch Kurven zu modellieren, ist die Kombination mehrerer Messungen. Jede Messung rekonstruiert jeweils lineare Bewegungen, aber über alle Messungen hinweg ergibt sich eine stückweise lineare Funktion, die sich der kurvenförmigen Bewegung annähert.

Als Zweites sei angemerkt, dass das nun vorgestellte Raum-Geschwindigkeits-Modell als großer Rahmen zu verstehen ist. An Details wie Detektorenabdeckungsbereiche und Messoperator werden keine Anforderungen gestellt. Die Gültigkeit des Modells ist daher auch für andere Anwendungen als der PET gegeben.

Zuletzt sei noch der Hinweis gegeben, dass im Folgenden unter dem Begriff „Zellen“ stets radioaktiv markierte Zellen gemeint sind.

Wir betrachten wie bisher ein sich über die Zeit veränderndes Maß  $\mu_t$ , das auf einem  $d$ -dimensionalen Hyperwürfel  $[0, 1]^d$  agiert. Es gilt also  $\mu_t \in M([0, 1]^d)$ , außerdem  $t \in [0, T]$  mit  $[0, T]$  erneut als Messzeitraum,  $T > 0$ . Das Maß beschreibt die Verteilung von insgesamt  $N \in \mathbb{N}$  Zellen, die sich allesamt innerhalb von  $[0, 1]^d$  befinden. Wie zuvor erwähnt, können wir mit einer möglichst kleinen Wahl von  $T$  annehmen, dass sich die Bewegungsgeschwindigkeit und Richtung der Zellen während des Messzeitraums nicht signifikant ändert. Demnach bewegt sich jede Zelle linear mit konstanter Geschwindigkeit.

Mathematisch modellieren wir die Position einer Zelle an der Stelle  $x \in [0, 1]^d$  durch ein Dirac-Maß

$$\delta_x(A) = \begin{cases} 1 & \text{falls } x \in A \\ 0 & \text{sonst} \end{cases},$$

für  $A \in \mathcal{B}([0, 1]^d)$ . Der obigen Annahme entsprechend können wir das Maß  $\mu_t$  nun ausformulieren als

$$\mu_t = \sum_{i=1}^N w_i \cdot \delta_{x_i + tv_i} \quad \text{mit } t \in [0, T],$$

wobei  $v_i \in \mathbb{R}^d$  den Geschwindigkeitsvektor der  $i$ -ten Zelle und  $x_i$  die Position der  $i$ -ten Zelle zum Startzeitpunkt  $t_0$  definieren. Durch den Term  $x_i + tv_i$  wird eine lineare Funktion mit  $t$  als Argument beschrieben, die genau der linearen Bewegung der  $i$ -ten Zelle im Gebiet  $[0, 1]^d$  entspricht. Die Intensität der Zellen (radioaktive Strahlung) beschreibt der Gewichtungsvektor  $w = (w_i)_i^N$ .

Insgesamt werden  $K \in \mathbb{N}$  Messungen zu den Zeitpunkten  $t_k := k \frac{T}{K}$  durchgeführt,  $k \in \{0, \dots, K-1\}$ , wobei ein Zeitpunkt stets als Startzeitpunkt einer Messung zu verstehen ist. Den Messzeitraum zum  $k$ -ten Zeitpunkt modellieren wir später mit  $[t_k, t_{k+1}]$ , woraus ersichtlich ist, dass  $t_k$  der Startzeit entspricht.

Zu jedem Zeitpunkt werden die Daten durch den gleichen (statischen) Messoperator erfasst:

$$\begin{aligned} M : M([0, 1]^d) &\longrightarrow \mathbb{R}^n \\ \mu_{t_k} &\longmapsto M(\mu_{t_k}) . \end{aligned} \tag{3.3}$$

Angewandt bedeutet diese Annahme, dass für jede der  $K$  Messungen der gleiche PET-Tomograph mit gleichem Messverfahren benutzt wird. Es entstehen folglich keine Fehler, die auf eine Veränderung der Messweise zurückzuführen sind. Die Daten der  $k$ -ten Messung sind von der Form

$$y_k := M(\mu_{t_k}) , \quad k \in \{0, \dots, K-1\} .$$

Zusammengefasst nehmen wir  $K$  Messungen von  $N$  Zellen mit linearer Bewegung in  $[0, 1]^d$ , mittels des Messoperator  $M$ , während der Zeit  $[0, T]$ , vor.

Bis hierhin besteht kein Unterschied zu einer möglichen statischen Rekonstruktion, die nun für jede der  $K$  Messungen die Position der Zellen ermitteln würde. Die Methode von Alberti et al. wird nachfolgend vorgestellt:

Die Geschwindigkeiten und Positionen einer jeden Zelle werden miteinander verknüpft, so dass

$$Z := \{(x_i, v_i) \mid i = 1, \dots, N\}$$

die Menge aller Zellenkonfigurationen erfasst. Jedes Element aus  $Z$  liegt nach Definition und obigen Annahmen im sogenannten *Phasenraum*:

$$\Omega := \left\{ (x, v) \in [0, 1]^d \times \mathbb{R}^d \mid x + tv \in [0, 1]^d , \text{ für alle } t \in [0, T] \right\} . \tag{3.4}$$

Hier kommt die Idee von Alberti et al. zur Geltung: Dadurch, dass der künftige Messoperator Maße auf dem Phasenraum als Argument benutzt und diese die Informationen über Geschwindigkeiten und Positionen aller Zellen zu jedem Zeitpunkt beinhalten, werden bei der Rekonstruktion dieser Maße nicht nur alle gemessenen Daten genutzt, sondern Positionen und Geschwindigkeiten auch simultan rekonstruiert.

Das Ziel besteht darin eine Verbindung zwischen Maßen auf  $[0, 1]^d$  und ihren sogenannten Repräsentanten auf  $\Omega$  herzuleiten, ferner einen Messoperator zu definieren, der für jedes Maß auf  $\Omega$  die zugehörigen Messdaten zu allen Zeitpunkten berechnet.

Für die Repräsentation von  $\mu_t$ ,  $t \in [0, T]$  fest, betrachten wir ein Maß  $\omega \in M(\Omega)$ , dessen

Form erneut einer Summe über Dirac-Maße entspricht. Wieder sei  $x_i \in [0, 1]^d$  die Position und  $v_i \in \mathbb{R}^d$  der Geschwindigkeitsvektor der  $i$ -ten Zelle. Mit  $\delta_{(x_i, v_i)} \in M(\Omega)$  beschreiben wir die Repräsentation der  $i$ -ten Zelle im Phasenraum und definieren

$$\omega = \sum_{i=1}^N w_i \delta_{(x_i, v_i)} .$$

Weiter definieren wir eine Funktion  $\phi$ , die die Verbindung zwischen Maßen auf  $[0, 1]^d$  und auf  $\Omega$  festhält, insbesondere also auch die Verbindung von  $\mu_t$  und seiner Repräsentation im Phasenraum. Für  $t \in [0, T]$  sei

$$\begin{aligned} \phi_t : \Omega &\longrightarrow [0, 1]^d \\ (x, v) &\longmapsto x + tv . \end{aligned}$$

Offensichtlich ist  $\phi_t$  surjektiv (setze  $v = 0$ ), aber nicht injektiv, da zum Beispiel für  $t = 0$   $\phi_0(0, 0) = \phi_0(0, 1)$  gilt. Ähnliche Gegenbeispiele lassen sich für alle  $t$  finden. Über das Bildmaß von  $\omega$  bezüglich  $\phi_t$  erhalten wir nun ein Maß auf  $[0, 1]^d$ :

$$(\phi_t)_\# \omega(\cdot) = \omega(\phi_t^{-1}(\cdot)) \in M([0, 1]^d) .$$

Sei  $A \subset [0, 1]^d$  beliebig, so gilt ausgeschrieben

$$(\phi_t)_\# \omega(A) = \omega(\phi_t^{-1}(A)) = \sum_{i=1}^N w_i \delta_{(x_i, v_i)}(\phi_t^{-1}(A)) .$$

Betrachten wir nun einen einzelnen Summanden ohne Gewichtung  $w_i$ :

$$\begin{aligned} \delta_{(x_i, v_i)}(\phi_t^{-1}(A)) &= \begin{cases} 1 & (x_i, v_i) \in \phi_t^{-1}(A) \\ 0 & \text{sonst} \end{cases} \\ &= \begin{cases} 1 & \phi_t(x_i, v_i) \in A \\ 0 & \text{sonst} \end{cases} \\ &= \begin{cases} 1 & x_i + tv_i \in A \\ 0 & \text{sonst} \end{cases} \\ &= \delta_{x_i + tv_i}(A) . \end{aligned}$$

Die zweite Gleichung folgt aus der Surjektivität von  $\phi_t$ . Dementsprechend gilt insgesamt

$$(\phi_t)_\# \omega(A) = \sum_{i=1}^N w_i \delta_{(x_i, v_i)}(\phi_t^{-1}(A)) = \sum_{i=1}^N w_i \delta_{x_i + tv_i}(A) = \mu_t(A) .$$

Da  $A$  beliebig gewählt war haben wir die Repräsentation des Maßes  $\mu_t$  im Phasenraum hergeleitet. Gemäß der Annahme, dass stets der (statische) Messoperator  $M$  die Messungen errechnet, besitzen die Daten für ein Maß  $\omega \in M(\Omega)$  zum Messzeitpunkt  $t_k$  die Form

$$\hat{y}_k := M((\phi_{t_k})_\# \omega) .$$

Als finalen Schritt konstruieren wir einen Messoperator, der alle Daten in einer  $n \times K$ -Matrix vereinigt. Auf triviale Weise werden die zu den  $K$  Zeitpunkten gemessenen Daten kombiniert, so dass die Funktionsvorschrift wie folgt definiert ist:

$$\begin{aligned} G : M(\Omega) &\longrightarrow \mathbb{R}^{n \times K} \\ \omega &\longmapsto \left( M((\phi_{t_k}) \# \omega) \right)_{k \in \{0, \dots, K-1\}} \\ &= (\hat{y}_k)_{k \in \{0, \dots, K-1\}} = G(\omega) =: y, \end{aligned} \quad (3.5)$$

In  $y$  als  $n \times K$ -Matrix werden alle zu jedem Zeitpunkt gemessenen Daten zusammengefasst. Den Messoperator  $G$  bezeichnen wir auch als *dynamischen Messoperator*. Nach der Durchführung eines kompletten Messvorgangs werden die Daten  $y$  erhalten und die Aufgabe besteht darin, das Maß  $\omega \in M(\Omega)$  zu rekonstruieren, gegeben  $y$ . Alberti et al. betrachten für die Lösung dieses Inversen Problems ein Minimierungsproblem, basierend auf der Minimierung bezüglich der TV-Norm,

$$\min_{\lambda \in M(\Omega)} \|\lambda\|_{TV}, \text{ so dass } G(\lambda) = y. \quad (3.6)$$

Da wir a-priori wissen, dass eine Linearkombination von Dirac-Maßen die Verteilung der Zellen beschreibt, rekonstruieren wir durch die Minimierung bezüglich der TV-Norm ein dünnbesetztes Maß aus  $M(\Omega)$  und demnach wieder eine Linearkombination von Dirac-Maßen (siehe Bemerkung 2.24).

### 3.4 Der dynamische Messoperator

In diesem Abschnitt wollen wir den Messoperator und insgesamt das zu lösende Minimierungsproblem herleiten. Wir kennen bereits den Vorwärtsoperator 3.2, und die Struktur des Messoperators 3.5.

Zunächst erfassen wir den Aufbau eines PET-Scanners mehr im Detail. Mit  $[0, 1]^d \subset \mathbb{R}^d$  beschreiben wir die Ausmaße des Tomographen. Insgesamt befinden sich  $D \in \mathbb{N}$  Detektoren am Rand von  $[0, 1]^d$ . Die Menge

$$D_{\text{paar}} := \left\{ (p, q) \mid p, q \in \{1, \dots, D\}, p > q \right\}$$

beinhaltet alle möglichen Paare von Detektoren, wobei ein Paar stets aus ungleichen Detektoren besteht. Weiter sei  $\Gamma_i \subset \partial[0, 1]^d$  das Gebiet des Randes des Tomographen, das durch den Detektor  $i \in \{1, \dots, D\}$  abgedeckt wird und für  $i \neq j$  soll  $\Gamma_i \cap \Gamma_j = \emptyset$  gelten. Durch das Bilden von

$$\Gamma_{(p,q)} := (\Gamma_p \times \Gamma_q) \cup (\Gamma_q \times \Gamma_p) \subset \partial[0, 1]^d \times \partial[0, 1]^d$$

besitzen wir eine Formulierung für das von dem Detektorenpaar  $(p, q) \in D_{\text{paar}}$  abgedeckte Gebiet. Mit

$$\Gamma := \bigcup_{(p,q) \in D_{\text{paar}}} \Gamma_{(p,q)} \subset \partial[0, 1]^d \times \partial[0, 1]^d$$

ist der komplette Abdeckungsbereich aller Detektorenpaare definiert. Da für die Detektoren

$$\bigcup_{i \in \{1, \dots, D\}} \Gamma_i \subsetneq \partial[0, 1]^d$$

gilt, folgt auch  $\Gamma \subsetneq \partial[0, 1]^d \times \partial[0, 1]^d$ . Dies ist die mathematische Formulierung dessen, dass zwischen benachbarten Detektoren nicht abgedeckte Zwischenräume existieren und wonach Zerfälle auch zu einer Nicht-Messung führen können.

Ähnlich zu 3.1 definieren wir durch

$$\mu = dt \times \frac{\ln 2}{T_{1/2}} \mu_t \in M([0, T] \times [0, 1]^d)$$

ein endliches Radonmaß. Erneut ist  $dt$  das auf dem Messzeitraum  $[0, T]$  definierte Lebesgue-Maß,  $T > 0$ , und  $\mu_t \in M([0, 1]^d)$  die Verteilung der  $N$  durch radioaktive Nuklide markierten Zellen zum Zeitpunkt  $t \in [0, T]$ . Gemäß Alberti gehen wir von folgender Struktur von  $\mu_t$  aus:

$$\mu_t = \sum_{i=1}^N \delta_{x_i + tv_i}, \quad t \in [0, T],$$

mit  $(x_i)_i \subset [0, 1]^d$  den Positionen und  $(v_i)_i \in \mathbb{R}^d$  den Geschwindigkeiten. Zwei Unterschiede bestehen jedoch: Zunächst ist die Intensität aller Nuklide standardisiert und wir haben  $(w_i)_i = (1)_i$  gesetzt. Als Zweites fällt der Faktor  $\frac{\ln 2}{T_{1/2}}$  auf. Er entspricht der Zerfallskonstanten der Nuklide und beschreibt die exponentielle Abnahme der Masse der Nuklide.  $T_{1/2}$  ist hier die Halbwertszeit des eingesetzten radioaktiven Materials.

Über den Messzeitraum werden  $K$  Messungen durchgeführt, wobei der Zeitraum der  $k$ -ten Messung gegeben ist durch

$$[t_k, t_{k+1}) \subset [0, T], \quad t_k := k \frac{T}{K} \text{ und } k \in \{0, \dots, K-1\}.$$

Dem Vorgehen von Alberti et al. folgend, gilt es zunächst, den statischen Messoperator

$$M : M([0, 1]^d) \longrightarrow \mathbb{R}^n,$$

der einem Maß auf  $[0, 1]^d$  einen  $n$ -dimensionalen Vektor zuordnet, herzuleiten. Für die Bestimmung der Dimension  $n$  erläutern wir nochmal die Messidee:

Pro Ereignis, dem Emittieren eines Positrons, entstehen zwei Photonen, die in verschiedene Richtungen fliegen. Beim Durchkreuzen des Randes von  $[0, 1]^d$  passieren beide im Messfall ein Gebiet  $\Gamma_{(p,q)}$  und werden vom Detektorenpaar  $(p, q)$  detektiert. Das Ereignis wollen wir demnach dem Detektorenpaar  $(p, q)$  zuschreiben. Angemerkt sei, dass für eine nicht gestreute Annihilierung der zu Grunde liegende Zerfall auf einer Linie zwischen den Detektoren des Paares liegt. Diese Linie wird auch *line of response* genannt.

Um nun den Zerfall abzuspeichern, indizieren wir zunächst die Menge  $D_{\text{paar}}$ . Mit der Gaußschen Summenformel werden alle möglichen Paare (auch gleiche wie  $(1, 1), \dots, (D, D)$ ) gezählt. Um die Gleichen herauszurechnen, subtrahieren wir  $D$ :

$$\left| D_{\text{paar}} \right| = -D + \sum_{i=1}^D i = \frac{D(D+1)}{2} - D = \frac{D(D-1)}{2} =: \widehat{D}$$

Demnach existiert für ein Detektorenpaar  $(p, q) \in D_{\text{paar}}$  ein eindeutiger Index  $i \in \{1, \dots, \widehat{D}\}$  und es gilt für die Dimension

$$n := \widehat{D} .$$

Der von Alberti et al. definierte Vektor speichert folglich die Anzahl an gemessenen Ereignissen eines jeden Detektorenpaares zeilenweise ab.

Das vom  $i$ -ten Detektorenpaar abgedeckte Gebiet setzen wir fest durch

$$\Gamma(i) := \Gamma_{(p,q)} ,$$

mit den zu  $i$  gehörenden Detektoren  $p$  und  $q$ .

Wie bereits erwähnt findet der Messprozess am Rand statt. Ein Maß der Form

$$dt \times F\left(\frac{\ln 2}{T_{1/2}} \mu_t\right) = dt \times \frac{\ln 2}{T_{1/2}} F(\mu_t) \in M([0, T] \times \partial[0, 1]^d \times \partial[0, 1]^d)$$

modelliert die durch den Vorwärtsoperator  $F$  (siehe (3.2)) transformierte Masse von  $\mu$  auf  $[0, T] \times \partial[0, 1]^d \times \partial[0, 1]^d$ . Aufgrund der Poisson-verteilten Wahrscheinlichkeit der Emittierung eines Positrons, entstehen die Messungen gemäß eines Poisson-Punkt-Prozesses (Definition 2.28) auf  $[0, T] \times \partial[0, 1]^d \times \partial[0, 1]^d$ . Diesen Poisson-Punkt-Prozesses gilt es im nächsten Schritt herzuleiten. Hierfür skalieren wir das Maß  $\mu_t$  mit  $\frac{T_{1/2}}{\ln 2}$ , um im Folgenden auf die Zerfallskonstante in der Notation zu verzichten.

Sei  $(\widetilde{\Omega}, \mathcal{F}, \mathbb{P})$  ein Wahrscheinlichkeitsraum,  $\lambda \in M([0, 1]^d)$  ein  $\sigma$ -endliches Maß und  $\eta_\lambda$  ein zu  $\lambda$  gehörender Punktprozess:

$$\begin{aligned} \eta_\lambda : (\widetilde{\Omega}, \mathcal{F}) &\longrightarrow \left( M([0, T] \times \partial[0, 1]^d \times \partial[0, 1]^d), \mathcal{N} \right) \\ \widetilde{\omega} &\longmapsto \eta_\lambda(\widetilde{\omega}, \cdot) =: \eta_\lambda(\widetilde{\omega})(\cdot) \end{aligned}$$

und

$$\mathcal{N} := \bigcup_{k \in \mathbb{N}_0} \left\{ \mu \in M([0, T] \times \partial[0, 1]^d \times \partial[0, 1]^d) \mid \mu(A) = k \right\}$$

$$A \in \mathcal{B}([0, T] \times \partial[0, 1]^d \times \partial[0, 1]^d)$$

die kleinste  $\sigma$ -Algebra, die für alle  $A \in \mathcal{B}([0, T] \times \partial[0, 1]^d \times \partial[0, 1]^d)$  die Abbildung  $\mu \longmapsto \mu(A)$  messbar macht. Ferner seien folgende Charakterisierungen gegeben:

- (i) Für alle  $A \in \mathcal{P}([0, T] \times \partial[0, 1]^d \times \partial[0, 1]^d)$  gilt:

$$\eta_\lambda(A) := \eta(\cdot)(A) \stackrel{d}{=} \text{Poi}(dt \times F(\lambda)(A)) .$$

- (ii) Für alle  $A_1, \dots, A_m \in \mathcal{P}([0, T] \times \partial[0, 1]^d \times \partial[0, 1]^d)$  mit  $A_i \cap A_j = \emptyset$  für  $i \neq j$  sind  $\eta_\lambda(A_1), \dots, \eta_\lambda(A_m)$  unabhängige Zufallsvariablen.

Durch diese Konstruktion erhalten wir für jedes  $\sigma$ -endliche Maß  $\lambda$  auf  $[0, 1]^d$  einen Poisson-Punkt-Prozess auf  $[0, T] \times \partial[0, 1]^d \times \partial[0, 1]^d$  mit Intensitätsmaß  $dt \times F(\lambda)$ . Mit Hilfe des Poisson-Punkt-Prozesses kann nun der statische Messoperator  $E_k : M([0, 1]^d) \rightarrow \mathbb{N}^{\widehat{D}}$  definiert werden. Zunächst werden dafür die Daten des  $i$ -ten Detektorpaares zur  $k$ -ten Messung konstruiert:

$$\begin{aligned} E_{k,i} : M([0, 1]^d) &\longrightarrow \{X : \widetilde{\Omega} \longrightarrow \mathbb{N} \mid X \text{ ist eine } \mathbb{N}\text{-wertige Zufallsvariable}\} \\ \mu_t &\longmapsto E_{k,i}(\mu_t) := \eta_{\mu_t}([t_k, t_{k+1}] \times \Gamma_{(i)}) \\ &\stackrel{d}{=} \text{Poi}\left((dt \times F(\mu_t))([t_k, t_{k+1}] \times \Gamma_{(i)})\right). \end{aligned}$$

Wir erhalten eine Funktion, die gemäß des Poisson-Punkt-Prozesses und des Vorwärtsoperators die Daten berechnet. An dieser Stelle sei darauf hingewiesen, dass mit  $E_{k,i}(\mu_{t_k})$  eine Poisson-verteilte Zufallsvariable vorliegt. Da für die Detektorenpaare  $i, j \in \{0, \dots, \widehat{D}\}$ , mit  $i \neq j$ ,  $\Gamma_{(i)} \cap \Gamma_{(j)} = \emptyset$  gilt, sind die Zufallsvariablen  $E_{k,i}(\mu_t)$  und  $E_{k,j}(\mu_t)$  nach obiger Charakterisierung unabhängig. Insgesamt gilt damit die paarweise Unabhängigkeit aller Zufallsvariablen. Außerdem betrachten wir nun abgeschlossene Zeitintervalle  $[t_k, t_{k+1}]$ , im Gegensatz zu der vorherigen Konstruktion über halboffene Intervalle. Dies macht in der Anwendung keinen Unterschied, da als Maß über die Zeit das Lebesgue-Maß gewählt wurde. Für den gesamten Messoperator zum  $k$ -ten Zeitpunkt gilt es nun diese Berechnungen über alle Detektorenpaare zu vereinen,

$$\begin{aligned} E_k : M([0, 1]^d) &\longrightarrow \{X : \widetilde{\Omega} \longrightarrow \mathbb{N} \mid X \text{ ist eine } \mathbb{N}\text{-wertige Zufallsvariable}\}^{\widehat{D}} \\ \mu_t &\longmapsto E_k(\mu_t) := \left(E_{k,i}(\mu_t)\right)_{i=1}^{\widehat{D}}. \end{aligned}$$

Für allgemeine  $\mu \in M([0, 1]^d)$  gilt aufgrund der Eigenschaften des Poisson-Punkt-Prozesses, dass  $E_l(\mu)$  unabhängig ist von  $E_j(\mu)$ , für alle  $l, j \in \{0, \dots, K-1\}$  mit  $l \neq j$ . Es gilt jedoch zu beachten, dass durch die Definition hin zu geschlossenen Intervallen  $[t_k, t_{k+1}]$  der Schnitt für beliebiges  $i$  und benachbarter  $l, j$  zwischen  $[t_l, t_{l+1}] \times \Gamma_{(i)} \cap [t_j, t_{j+1}] \times \Gamma_{(i)}$  nicht leer ist. Dennoch ist die Unabhängigkeit gegeben, folgernd aus den Eigenschaften des Lebesgue-Maß (der Schnitt ist eine Nullmenge). Der Messprozess ist insgesamt zu jedem Zeitpunkt der Gleiche und erfüllt die Bedingung an den statischen Messoperator (3.3).

Wir wechseln auf Maße die auf dem Phasenraum

$$\Omega = \left\{ (x, v) \in [0, 1]^d \times \mathbb{R}^d \mid x + tv \in [0, 1]^d, \text{ für alle } t \in [0, T] \right\}$$

agieren. Wie in Abschnitt 3.3 definieren wir über

$$\begin{aligned} \phi_t : \Omega &\longrightarrow [0, 1]^d \\ (x, v) &\longmapsto x + tv, \end{aligned}$$

und

$$\omega = \sum_{i=1}^N \delta_{(x_i, v_i)} \in M(\Omega) ,$$

die Repräsentation von  $\mu_t$  mit  $(\phi_t)_\# \omega = \mu_t$ . Gemäß Alberti et al. konstruieren wir  $E_k$  zu einem Operator um, der als Argument Maße aus  $M(\Omega)$  benutzt. Dafür greifen wir auf jene Repräsentation von Maßen zurück, nutzen folglich das Bildmaß von Maßen auf  $[0, 1]^d$  unter  $\phi_t$ . Über den Vorwärtsoperator erhalten wir die im Grunde identisch gebliebene Struktur des Messoperators, nur mit Änderung der Definitionsmenge:

$$\begin{aligned} \widehat{E}_{k,i} : M(\Omega) &\longrightarrow \{X : \widetilde{\Omega} \longrightarrow \mathbb{N} \mid X \text{ ist eine } \mathbb{N}\text{-wertige Zufallsvariable}\} \\ \omega &\longmapsto \widehat{E}_{k,i}(\omega) := E_{k,i}((\phi_t)_\# \omega) \\ &\stackrel{d}{=} \text{Poi}\left(dt \times F((\phi_t)_\# \omega)([t_k, t_{k+1}] \times \Gamma_{(i)})\right) . \end{aligned}$$

Es gilt erneut die einzelnen berechneten Detektoren-Daten zusammenzufassen, via

$$\begin{aligned} \widehat{E}_k : M(\Omega) &\longrightarrow \{X : \widetilde{\Omega} \longrightarrow \mathbb{N} \mid X \text{ ist eine } \mathbb{N}\text{-wertige Zufallsvariable}\}^{\widehat{D}} \\ \omega &\longmapsto \widehat{E}_k(\omega) := \left(\widehat{E}_{k,i}(\omega)\right)_{i=1}^{\widehat{D}} . \end{aligned}$$

$\widehat{E}_k$  entspricht der statischen Messung zum Zeitpunkt  $t_k$ ,  $k \in \{0, \dots, K-1\}$ . Der letzte Schritt hin zu der Aufstellung des dynamischen Messoperators nach 3.5 ist ähnlich und besteht daraus, alle  $K$  Messungen in einer Matrix zusammenzuführen.

$$\begin{aligned} G : M(\Omega) &\longrightarrow \{X : \widetilde{\Omega} \longrightarrow \mathbb{N} \mid X \text{ ist eine } \mathbb{N}\text{-wertige Zufallsvariable}\}^{\widehat{D} \times K} \\ \omega &\longmapsto Y := G(\omega) = \left(\widehat{E}_k(\omega)\right)_{k=0}^{K-1} . \end{aligned} \tag{3.7}$$

Die gemessenen Daten entsprechen somit einer Poisson-verteilten Zufallsmatrix, deren Einträge paarweise unabhängige Zufallsvariablen sind:

$$Y \stackrel{d}{=} \left( \left( \text{Poi}\left(dt \times F((\phi_t)_\# \omega)([t_k, t_{k+1}] \times \Gamma_{(i)})\right) \right)_{i=1}^{\widehat{D}} \right)_{k=0}^{K-1} .$$

Wenn eine Messung durch einen PET-Tomographen vorgenommen wird, so werden die Messungen Realisierungen  $y$  von  $Y$  sein. Aufgrund dieser Eigenschaft bietet sich für die Lösung des von Alberti et al. vorgeschlagenen Minimierungsproblems

$$\min_{\lambda \in M(\Omega)} \|\lambda\|_{TV} , \text{ so dass } G(\lambda)(\widetilde{\omega}) = y , \tag{3.8}$$

eine stochastische Modellierung an. Damit kann die Rekonstruktion der den Daten  $y$  zugrundeliegenden Verteilung von Zellen erreicht werden. Hier ist  $\widetilde{\omega}$  Element des Ereignisraums  $\widetilde{\Omega}$

und soll verdeutlichen, dass  $y$  eine Realisierung ist. Veranschaulicht entscheidet  $\tilde{\omega}$  die Richtlinie, nach der sich die radioaktiven Zellen verhalten, sprich zu welchem Zeitpunkt und an welchem Ort, ein Positron emittiert wird. Würden wir unser Vorhaben bis zum kleinsten Detail durchdenken, so sind wir eigentlich auf der Suche nach genau jenem  $\tilde{\omega}$ , schließlich ist dieses der grundlegendste Auslöser der zu Messungen führt. Jedoch ist diese Ermittlung von theoretischer Natur und für unsere Zwecke nicht zielführend. Bis heute besteht der allgemeine Konsens, dass hinter dem natürlichen Zerfall eines radioaktiven Nuklids reiner Zufall steckt. Unser Ziel bleibt die Rekonstruktion des Ortes des radioaktiven Zerfalls und daraus folgernd die Ermittlung der Bewegungen der Zellen.



## 4 Maximum-a-posteriori-Schätzung und das Rekonstruktionsfunktional

Das Ziel ist also, durch das Lösen des Minimierungsproblems 3.8 mit zugehörigem dynamischen Operator 3.7, das Maß

$$\hat{\lambda} = \sum_{i=1}^N \delta_{(x_i, v_i)} \in M(\Omega) \quad (4.1)$$

zu rekonstruieren. Hier ist  $\hat{\lambda}$  die Repräsentation des Maßes

$$\hat{\mu}_t = \sum_{i=1}^N \delta_{x_i + tv_i} \in M([0, 1]^d)$$

im Phasenraum  $\Omega$  mit dem Zusammenhang  $(\phi_t)_\# \lambda = \mu_t$ .

### 4.1 Eine erste Variante des Rekonstruktionsfunktionals

Gehen wir davon aus, dass wir einen Datensatz  $y$  gemessen haben. Gemäß der Struktur des Operators  $G$  sind die Daten  $y = (y_{k,i})_{k,i}$  eine Realisierung der Zufallsmatrix  $G(\hat{\lambda})$  und wir notieren mit  $y_{k,i}$  den  $k, i$ -ten Eintrag. Dieser entspricht der gemessenen Intensität des  $i$ -ten Detektorenpaars zur  $k$ -ten Messung. Ferner notieren wir für Auswertungen des Intensitätsmaßes des zugehörigen Poisson-Punkt-Prozesses

$$C_{k,i}(\lambda) := \left( dt \times F((\phi_t)_\# \lambda) ([t_k, t_{k+1}] \times \Gamma_{(i)}) \right) \quad (= \mathbb{E}_\lambda Y_{k,i}) .$$

Hier kommt Campbells Formel (Satz 2.27) versteckt zum Einsatz, um den Erwartungswert des Poisson-Punkt-Prozess auszuwerten. Analog zu oben gilt

$$Y = (Y_{k,i})_{k,i} = G(\hat{\lambda}) \stackrel{d}{=} \left( \text{Poi}(C_{k,i}(\hat{\lambda})) \right)_{k,i} .$$

Für die Lösung des Minimierungsproblems 3.8 benutzen wir einen bayesianischen Ansatz und den MAP-Schätzer nach Definition 2.22, um die Schlechtgestellttheit des Problems zu lösen. In diesem Kontext entspricht der MAP-Schätzer

$$\lambda^* = \underset{\lambda \in M(\Omega)}{\text{argmin}} - \log \mathbb{P}[Y = y | \lambda] - \log \mathbb{P}[\lambda] , \quad (4.2)$$

wonach wir im nächsten Schritt die einzelnen Wahrscheinlichkeiten bestimmen oder festlegen müssen. Da unser Minimierungsproblem Maße mit niedriger TV-Norm sucht, modellieren wir die Wahrscheinlichkeit, dass das Maß  $\lambda$  vorliegt, über den Gibbs-Prior mit der TV-Norm als Regularisierungsterm. Wie in Bemerkung 2.24 bereits erwähnt, gilt

$$\begin{aligned}\mathbb{P}[\lambda] &= \exp(-\tilde{\beta}\|\lambda\|_{TV}), \text{ bzw.} \\ \log \mathbb{P}[\lambda] &= -\tilde{\beta}\|\lambda\|_{TV}.\end{aligned}\tag{4.3}$$

Der Regularisierungsparameter  $\tilde{\beta}$  ist hier größer als null zu wählen und steuert den Einfluss der Regularisierung. Da wir in Abschnitt 3.4 das Maß mit dem Inversen der zu den Nukliden gehörenden Zerfallskonstanten multipliziert haben, setzen wir im Folgenden  $\beta := \tilde{\beta} \cdot \frac{\ln 2}{T_{1/2}}$ . Je höher die zu einem Maß gehörende TV-Norm ist, umso geringer ist die Wahrscheinlichkeit, dass das Maß vorliegt. Dies entspricht dem Minimierungsproblem, schließlich wird über die TV-Norm minimiert. Es bleibt noch die bedingte Wahrscheinlichkeit von  $Y = y$ , bezüglich  $\lambda$  zu berechnen. Für diese Wahrscheinlichkeit gilt

$$\begin{aligned}\mathbb{P}[Y = y|\lambda] &= \mathbb{P}[(Y_{k,i})_{k,i} = (y_{k,i})_{k,i}|\lambda] \\ &= \prod_{k=0}^{K-1} \prod_{i=1}^{\hat{D}} \mathbb{P}_{C_{k,i}(\lambda)}[Y_{k,i} = y_{k,i}] \\ &= \prod_{k=0}^{K-1} \prod_{i=1}^{\hat{D}} \frac{C_{k,i}(\lambda)^{y_{k,i}}}{y_{k,i}!} e^{-C_{k,i}(\lambda)}.\end{aligned}$$

Die Gleichheit ergibt sich zum einen aus der Unabhängigkeit der Zufallsvariablen  $(Y_{k,i})_{k,i}$  (Unabhängigkeit von Poisson-Punkt-Prozessen auf disjunkten Mengen). Zum anderen sind die Zufallsvariablen  $Y_{k,i}$ , bedingt durch das Vorliegen von  $\lambda$ ,  $\text{Poi}(C_{k,i}(\lambda))$ -verteilt, was durch die Indizierung des Wahrscheinlichkeitsmaßes  $\mathbb{P}$ , via  $\mathbb{P}_{C_{k,i}(\lambda)}$ , ausgedrückt wird. Alle nötigen Wahrscheinlichkeiten sind ermittelt und wir können als nächsten Schritt die dem MAP-Schätzer zu Grunde liegende Minimierungsfunktion herleiten:

$$\begin{aligned}-\log \mathbb{P}[Y = y|\lambda] - \log \mathbb{P}[\lambda] &= -\log \left( \prod_{k=0}^{K-1} \prod_{i=1}^{\hat{D}} \frac{C_{k,i}(\lambda)^{y_{k,i}}}{y_{k,i}!} e^{-C_{k,i}(\lambda)} \right) - (-\beta\|\lambda\|_{TV}) \\ &= -\sum_{k=0}^{K-1} \sum_{i=1}^{\hat{D}} \log \left( \frac{1}{y_{k,i}!} C_{k,i}(\lambda)^{y_{k,i}} e^{-C_{k,i}(\lambda)} \right) + \beta\|\lambda\|_{TV} \\ &= -\sum_{k=0}^{K-1} \sum_{i=1}^{\hat{D}} -\log y_{k,i}! + y_{k,i} \log C_{k,i}(\lambda) - C_{k,i}(\lambda) + \beta\|\lambda\|_{TV} \\ &= \sum_{k=0}^{K-1} \sum_{i=1}^{\hat{D}} \log y_{k,i}! + \sum_{k=0}^{K-1} \sum_{i=1}^{\hat{D}} C_{k,i}(\lambda) - y_{k,i} \log C_{k,i}(\lambda) + \beta\|\lambda\|_{TV}.\end{aligned}$$

Da über  $\lambda \in M(\Omega)$  minimiert wird, können wir den ersten Doppelsummanden-Term über  $\log y_{k,i}!$  außen vor lassen und erhalten insgesamt eine erste Formulierung für ein zu mini-

mierendes Funktional:

$$\widehat{J}^y(\lambda) := \sum_{k=0}^{K-1} \sum_{i=1}^{\widehat{D}} C_{k,i}(\lambda) - y_{k,i} \log C_{k,i}(\lambda) + \beta \|\lambda\|_{TV} . \quad (4.4)$$

Wir rekapitulieren an dieser Stelle noch einmal und erklären die Notationen. Mit  $y_{k,i}$  wird die durch das Detektorenpaar  $i$  zum Messzeitpunkt  $k$  gemessene Anzahl an Ereignissen beschrieben, während  $C_{k,i}(\lambda)$  die Auswertung des Intensitätsmaßes des Poisson-Punkt-Prozesses für das  $i$ -te Detektorenpaar zur  $k$ -ten Messung beim Vorliegen von  $\lambda$  erfasst. In diesem Kontext entspricht  $C_{k,i}$  dem Erwartungswert einer  $\text{Poi}(C_{k,i}(\lambda))$  verteilten Zufallsvariable, also der *erwarteten Anzahl an Messungen* des Detektorenpaares  $i$  zum Messzeitpunkt  $k$ .

Unsere Funktion 4.4 sucht demnach ein Maß, dessen erwartete Messungen den gemessenen Daten weitestgehend entsprechen. In der Tat erkennen wir unter Weglassung des Regularisierungsterms  $\|\lambda\|_{TV}$  und nach einfachen Rechnungen, dass das Minimum von  $\widehat{J}^y(\lambda)$  erreicht wird, wenn  $C_{k,i}(\lambda) = y_{k,i}$  für alle  $k, i$  gilt.

## 4.2 Problematik und Unterscheidung der Daten

Leider scheidert jedoch unsere Funktion in der Anwendung, das Maß  $\widehat{\lambda}$  zu rekonstruieren. Der Grund hierfür liegt an der starken Tendenz dazu, gemessene Ereignisse stets als ungestreute Photonenstrahlung zu interpretieren. Um dies zu sehen, betrachten wir für ein festes  $k$  und  $i$  den Term  $C_{k,i}(\lambda)$  genauer:

$$\begin{aligned} C_{k,i}(\lambda) &= dt \times F((\phi_t)_{\#}\lambda)([t_k, t_{k+1}] \times \Gamma_{(i)}) \\ &= \int_{t_k}^{t_{k+1}} F((\phi_t)_{\#}\lambda)(\Gamma_{(i)}) dt . \end{aligned}$$

Wie wir in der Herleitung des Vorwärtsoperators gezeigt haben, ist dieser eine Linearkombination zweier Operatoren. Einer modelliert die Messung von gestreuten Photonen, der andere die Messung von genau in entgegengesetzte Richtung fliegenden Photonen. Dabei wird die Intensität der gestreuten Photonen gleichmäßig auf alle Detektoren verteilt und ist für einen einzelnen Detektor kleiner als die gebündelte Intensität eines ungestreuten Ereignisses. Diese Eigenschaft führt bei der Minimierung von  $\widehat{J}^y(\lambda)$  zu dem oben beschriebenen Phänomen, die gemessenen Ereignisse als nicht-gestreute Ereignisse zu identifizieren. Ein einzelner Summand von  $\widehat{J}^y$  besitzt die Form

$$C_{k,i}(\lambda) - y_{k,i} \log C_{k,i}(\lambda) ,$$

und gerade kleine Intensitäten (unter dem Wert 1) werden hier stark bestraft, schließlich betrachten wir den negativen Logarithmus, der gegen unendlich läuft für Werte nahe 0. Gilt nun  $y_{k,i} \geq 1$ , so wird bei der Minimierung  $C_{k,i}(\lambda) \geq 1$  gesucht. Demnach wird die gemessene Intensität stets auf ein Detektorenpaar konzentriert, anstatt diese gleichmäßig auf alle Detektorenpaare zu verteilen und gestreute Ereignisse sind wesentlich benachteiligt.

Das Minimierungsfunktional 4.4 muss folglich umgeformt werden, um eine geeignetere Variante zu erhalten. Hierfür unterteilen wir die Zufallsmatrix  $Y$  in zwei Zufallsmatrizen  $Y^E$  und  $Y^S$ , die die nicht-gestreuten und gestreuten Messungen separat erfassen:

$$\left(Y_{k,i}\right)_{k,i} = \left(Y_{k,i}^E\right)_{k,i} + \left(Y_{k,i}^S\right)_{k,i} .$$

Einen gemessenen Datensatz unterteilen wir ebenfalls in durch gestreute, bzw. perfekt fliegende Photonen ausgelöste Ereignisse,

$$\left(y_{k,i}\right)_{k,i} = \left(y_{k,i}^E\right)_{k,i} + \left(y_{k,i}^S\right)_{k,i} .$$

Diesen Ansatz führen wir erneut für den Term  $C_{k,i}(\cdot)$  durch, unterteilen also auch hier die Intensität in die durch gestreute Ereignisse und ungestreute Ereignisse gemessene Intensität. Dabei nutzen wir aus, dass für den Vorwärtsoperator  $F = p_S F^S + p_E F^E$  gilt, siehe 3.2, und erhalten

$$\begin{aligned} C_{k,i}(\lambda) &= dt \times F((\phi_t)_{\#} \lambda)([t_k, t_{k+1}] \times \Gamma_{(i)}) \\ &= \int_{t_k}^{t_{k+1}} (p_S F^S + p_E F^E)((\phi_t)_{\#} \lambda)(\Gamma_{(i)}) dt \\ &= \underbrace{\int_{t_k}^{t_{k+1}} p_S F^S((\phi_t)_{\#} \lambda)(\Gamma_{(i)}) dt}_{=: C_{k,i}^S(\lambda)} + \underbrace{\int_{t_k}^{t_{k+1}} p_E F^E((\phi_t)_{\#} \lambda)(\Gamma_{(i)}) dt}_{=: C_{k,i}^E(\lambda)} \\ &= C_{k,i}^S(\lambda) + C_{k,i}^E(\lambda) . \end{aligned}$$

Unter der Annahme den gestreuten Anteil der Messungen zu kennen, modelliert durch

$$|y^S|_1 = \sum_{k=0}^{K-1} \sum_{i=1}^{\hat{D}} |y_{k,i}^S| = b , \text{ für ein bekanntes } b \in \mathbb{N}_0 ,$$

stellen wir ein neues Minimierungsproblem auf:

$$\begin{aligned} \min_{\lambda \in M(\Omega)} \|\lambda\|_{TV} , \text{ so dass } G(\lambda) = y^E + y^S = y \\ \text{mit } |y^S|_1 = b \end{aligned} \quad . \quad (4.5)$$

Erneut wenden wir die MAP-Schätzung an und leiten das zugehörige Minimierungsfunktional her. Die zu maximierende bedingte Wahrscheinlichkeit ist nach dem Satz von Bayes von der Form:

$$\mathbb{P}[\lambda | Y^E = y^E, Y^S = y^S] = \frac{\mathbb{P}[Y^E = y^E, Y^S = y^S | \lambda] \mathbb{P}[\lambda]}{\mathbb{P}[Y^E = y^E, Y^S = y^S]}$$

und der MAP-Schätzer ist gegeben durch:

$$\lambda^* = \operatorname{argmin}_{\lambda \in M(\Omega)} -\log \mathbb{P}[Y^E = y^E, Y^S = y^S | \lambda] - \log \mathbb{P}[\lambda] .$$

Der Unterschied zu 4.2 besteht also in der bedingten Wahrscheinlichkeit. Mit Nutzung der Unabhängigkeit aller Zufallsvariablen  $Y_{k,i}^E$  und  $Y_{k,i}^S$  können wir diese analog umformen:

$$\begin{aligned}
 -\log \mathbb{P}[Y^E = y^E, Y^S = y^S | \lambda] &= -\log \left( \prod_{k=0}^{K-1} \prod_{i=1}^{\hat{D}} \mathbb{P}[Y_{k,i}^E = y_{k,i}^E, Y_{k,i}^S = y_{k,i}^S] \right) \\
 &= -\log \left( \prod_{k=0}^{K-1} \prod_{i=1}^{\hat{D}} \mathbb{P}_{C_{k,i}^E(\lambda)}[Y_{k,i}^E = y_{k,i}^E] \cdot \mathbb{P}_{C_{k,i}^S(\lambda)}[Y_{k,i}^S = y_{k,i}^S] \right) \\
 &= -\log \left( \prod_{k=0}^{K-1} \prod_{i=1}^{\hat{D}} \frac{C_{k,i}^E(\lambda)^{y_{k,i}^E} \cdot C_{k,i}^S(\lambda)^{y_{k,i}^S}}{y_{k,i}^E! \cdot y_{k,i}^S!} \cdot e^{-C_{k,i}(\lambda)} \right) \\
 &= -\sum_{k=0}^{K-1} \sum_{i=1}^{\hat{D}} y_{k,i}^E \log C_{k,i}^E(\lambda) + y_{k,i}^S \log C_{k,i}^S(\lambda) - C_{k,i}(\lambda) - \log y_{k,i}^E! - \log y_{k,i}^S! .
 \end{aligned}$$

Ebenfalls mit dem Gibbs-Prior (4.3) und dem Weglassen der Konstanten  $\log y_{k,i}^E!$  und  $\log y_{k,i}^S!$  erhalten wir das Minimierungsfunktional für unser Minimierungsproblem als

$$\tilde{J}^{y^E, y^S}(\lambda) := \sum_{k=0}^{K-1} \sum_{i=1}^{\hat{D}} C_{k,i}(\lambda) - y_{k,i}^E \log C_{k,i}^E(\lambda) - y_{k,i}^S \log C_{k,i}^S(\lambda) + \beta \|\lambda\|_{TV} .$$

Der MAP-Schätzer ist folglich

$$\lambda^* = \underset{\substack{\lambda \in M(\Omega) \\ \text{mit } |y^S|_1 = b \\ \text{und } y^E + y^S = y}}{\operatorname{argmin}} \tilde{J}^{y^E, y^S}(\lambda) .$$

Wir erhalten ein Minimierungsfunktional, dem genau vorgesagt wird, wie es die Anzahl an gestreuten Ereignissen zu wählen hat. Somit ist garantiert, dass bei der Minimierung das Verhältnis zwischen Streuung und Nicht-Streuung stimmt, anders als bei  $\hat{J}^y(\lambda)$ .

Auch wenn das Funktional uns eine geeignete Lösung liefert, so besitzt es einen großen Nachteil: Die Minimierung mit der Nebenbedingung  $y^S + y^E = y$  ist aufgrund ihres kombinatorischen Charakters numerisch schwer umzusetzen. Ein Algorithmus würde sehr lange brauchen, bis der MAP-Schätzer ermittelt wird, eine in der Anwendung nicht wünschenswerte Eigenschaft von Algorithmen. Erneut schafft eine Umformulierung von  $\tilde{J}^{y^E, y^S}(\lambda)$  Abhilfe.

### 4.3 Gewichtetes Maximums-Funktional

Wir betrachten das gewichtete Maximums-Funktional, dessen Minimierer ebenfalls ein Minimierer von  $\tilde{J}^{y^E, y^S}(\lambda)$  sein soll. Hierfür führen wir einen Streuungsparameter  $p > 0$  ein,

der den Einfluss von  $C_{k,i}^S(\lambda)$  bestimmt. Der Parameter  $p$  ist wie folgt zu wählen: Es soll gelten:

$$\sum_{k=0}^{K-1} \sum_{i=1}^{\widehat{D}} y_{k,i}^S \log(C_{k,i}^E(\lambda)) \leq \sum_{k=0}^{K-1} \sum_{i=1}^{\widehat{D}} y_{k,i}^S \log(p \cdot C_{k,i}^S(\lambda)), \text{ falls } |Y^E| \leq |Y^S|. \quad (4.6)$$

Andernfalls, wenn für Messungen primär nicht-gestreute Photonen verantwortlich sind, soll die Umkehrung für jene Messereignisse  $y_{k,i}^E$  gelten:

$$\sum_{k=0}^{K-1} \sum_{i=1}^{\widehat{D}} y_{k,i}^E \log(C_{k,i}^E(\lambda)) \geq \sum_{k=0}^{K-1} \sum_{i=1}^{\widehat{D}} y_{k,i}^E \log(p \cdot C_{k,i}^S(\lambda)), \text{ falls } |Y^E| \geq |Y^S|. \quad (4.7)$$

Wir haben mit  $p$  nach Konstruktion ein Werkzeug, das die Werte der Intensitätseinflüsse anpasst. Das gewichtete Maximums-Funktional ist nun definiert durch

$$\widehat{J}^{y,p}(\lambda) := \sum_{k=0}^{K-1} \sum_{i=1}^{\widehat{D}} C_{k,i}(\lambda) - y_{k,i} \log(\max(p \cdot C_{k,i}^S(\lambda), C_{k,i}^E(\lambda))) + \beta \|\lambda\|_{TV}. \quad (4.8)$$

Es gilt mit geeigneter Wahl von  $p$  zu zeigen, dass ein Minimierer von  $\widehat{J}^{y,p}$  ebenfalls das Minimierungsproblem 4.5 löst, also ein Minimierer von  $\widetilde{J}^{y^E, y^S}$  ist. Hierfür sei  $p$  wie zuvor beschrieben gewählt. Weiter sei mit  $\widehat{\lambda} \in M(\Omega)$  ein Minimierer von  $\widehat{J}^{y,p}$  gegeben und  $\widetilde{\lambda} \in M(\Omega)$  sei ein Minimierer von  $\widetilde{J}^{y^E, y^S}$ , der mit zugehöriger Aufteilung der Messungen  $\widetilde{y}^E + \widetilde{y}^S = y$  auch die Nebenbedingung  $|\widetilde{y}^S|_1 = b$  erfüllt. Wir setzen in das Funktional  $\widetilde{J}^{y^E, y^S}$  den Minimierer von  $\widehat{J}^{y,p}$  als Argument ein und schätzen ab:

$$\begin{aligned} \widetilde{J}^{y^E, \widetilde{y}^S}(\widetilde{\lambda}) &\leq \widetilde{J}^{y^E, \widetilde{y}^S}(\widehat{\lambda}) = \sum_{k=0}^{K-1} \sum_{i=1}^{\widehat{D}} C_{k,i}(\widehat{\lambda}) - \widetilde{y}_{k,i}^E \log C_{k,i}^E(\widehat{\lambda}) - \widetilde{y}_{k,i}^S \log C_{k,i}^S(\widehat{\lambda}) + \beta \|\widehat{\lambda}\|_{TV} \\ &= \sum_{k=0}^{K-1} \sum_{i=1}^{\widehat{D}} C_{k,i}(\widehat{\lambda}) - \widetilde{y}_{k,i}^E \log C_{k,i}^E(\widehat{\lambda}) - \widetilde{y}_{k,i}^S \log C_{k,i}^S(\widehat{\lambda}) \\ &\quad - \underbrace{\left( \sum_{k=0}^{K-1} \sum_{i=1}^{\widehat{D}} \widetilde{y}_{k,i}^S \log p - b \log p \right)}_{=0} + \beta \|\widehat{\lambda}\|_{TV} \\ &= \left( \sum_{k=0}^{K-1} \sum_{i=1}^{\widehat{D}} C_{k,i}(\widehat{\lambda}) - \widetilde{y}_{k,i}^E \log C_{k,i}^E(\widehat{\lambda}) - \widetilde{y}_{k,i}^S \log(p \cdot C_{k,i}^S(\widehat{\lambda})) \right) + b \log p + \beta \|\widehat{\lambda}\|_{TV}. \end{aligned}$$

Durch die Trennung der Messung  $y_{k,i} = \widetilde{y}_{k,i}^E + \widetilde{y}_{k,i}^S$  können wir folgende Fallunterscheidung treffen: Zunächst nehmen wir  $|Y^E| \leq |Y^S|$  an. Es folgt mit den Annahmen an  $p$

$$\begin{aligned} - \sum_{k=0}^{K-1} \sum_{i=1}^{\widehat{D}} \widetilde{y}_{k,i}^E \log(C_{k,i}^E(\widehat{\lambda})) - \underbrace{\sum_{k=0}^{K-1} \sum_{i=1}^{\widehat{D}} \widetilde{y}_{k,i}^S \log(p \cdot C_{k,i}^S(\widehat{\lambda}))}_{\geq \sum_{k=0}^{K-1} \sum_{i=1}^{\widehat{D}} \widetilde{y}_{k,i}^S \log(C_{k,i}^E(\widehat{\lambda}))} &\leq - \sum_{k=0}^{K-1} \sum_{i=1}^{\widehat{D}} y_{k,i} \log(C_{k,i}^E(\widehat{\lambda})), \end{aligned}$$

und für  $|Y^E| \geq |Y^S|$  gilt:

$$\begin{aligned} & \underbrace{\sum_{k=0}^{K-1} \sum_{i=1}^{\hat{D}} -\tilde{y}_{k,i}^E \log(C_{k,i}^E(\hat{\lambda}))}_{\leq \sum_{k=0}^{K-1} \sum_{i=1}^{\hat{D}} \tilde{y}_{k,i}^E \log(p \cdot C_{k,i}^S(\lambda))} - \sum_{k=0}^{K-1} \sum_{i=1}^{\hat{D}} \tilde{y}_{k,i}^S \log(p \cdot C_{k,i}^S(\hat{\lambda})) \leq - \sum_{k=0}^{K-1} \sum_{i=1}^{\hat{D}} y_{k,i} \log(p \cdot C_{k,i}^S(\hat{\lambda})) . \end{aligned}$$

Somit erhalten wir insgesamt:

$$\begin{aligned} & \sum_{k=0}^{K-1} \sum_{i=1}^{\hat{D}} -\tilde{y}_{k,i}^E \log(C_{k,i}^E(\hat{\lambda})) - \tilde{y}_{k,i}^S \log(p \cdot C_{k,i}^S(\hat{\lambda})) \\ & \leq \sum_{k=0}^{K-1} \sum_{i=1}^{\hat{D}} -y_{k,i} \log\left(\max(p \cdot C_{k,i}^S(\hat{\lambda}), C_{k,i}^E(\hat{\lambda}))\right) . \end{aligned}$$

Unter Hinzunahme der Ungleichung können wir mit der Abschätzung fortfahren.

$$\begin{aligned} \tilde{J}^{\tilde{y}^E, \tilde{y}^S}(\hat{\lambda}) &= \left( \sum_{k=0}^{K-1} \sum_{i=1}^{\hat{D}} C_{k,i}(\hat{\lambda}) - \tilde{y}_{k,i}^E \log C_{k,i}^E(\hat{\lambda}) - \tilde{y}_{k,i}^S \log(p \cdot C_{k,i}^S(\hat{\lambda})) \right) + b \log p + \beta \|\hat{\lambda}\|_{TV} \\ &\leq \left( \sum_{k=0}^{K-1} \sum_{i=1}^{\hat{D}} C_{k,i}(\hat{\lambda}) - y_{k,i} \log\left(\max(p \cdot C_{k,i}^S(\hat{\lambda}), C_{k,i}^E(\hat{\lambda}))\right) \right) + b \log p + \beta \|\hat{\lambda}\|_{TV} \\ &= \hat{J}^{y,p}(\hat{\lambda}) + b \log p . \end{aligned}$$

Da nach Konstruktion und Annahme  $\hat{\lambda}$  Minimierer von  $\hat{J}^{y,p}$  ist, gilt  $\hat{J}^{y,p}(\hat{\lambda}) \leq \hat{J}^{y,p}(\lambda)$  für alle  $\lambda \in M(\Omega)$ , insbesondere also auch für  $\tilde{\lambda}$ , den Minimierer von  $\tilde{J}^{\tilde{y}^E, \tilde{y}^S}$ :

$$\begin{aligned} \hat{J}^{y,p}(\hat{\lambda}) + b \log p &\leq \hat{J}^{y,p}(\tilde{\lambda}) + b \log p = \hat{J}^{\tilde{y}^E + \tilde{y}^S, p}(\tilde{\lambda}) + b \log p \\ &= \left( \sum_{k=0}^{K-1} \sum_{i=1}^{\hat{D}} C_{k,i}(\tilde{\lambda}) - \tilde{y}_{k,i} \log\left(\max(p \cdot C_{k,i}^S(\tilde{\lambda}), C_{k,i}^E(\tilde{\lambda}))\right) \right) + b \log p + \beta \|\tilde{\lambda}\|_{TV} \\ &= \left( \sum_{k=0}^{K-1} \sum_{i=1}^{\hat{D}} C_{k,i}(\tilde{\lambda}) - \underbrace{\tilde{y}_{k,i}^E \log\left(\max(p \cdot C_{k,i}^S(\tilde{\lambda}), C_{k,i}^E(\tilde{\lambda}))\right)}_{\geq \log C_{k,i}^E(\tilde{\lambda})} \right. \\ &\quad \left. - \underbrace{\tilde{y}_{k,i}^S \log\left(\max(p \cdot C_{k,i}^S(\tilde{\lambda}), C_{k,i}^E(\tilde{\lambda}))\right)}_{\geq \log(p \cdot C_{k,i}^S(\tilde{\lambda}))} \right) + b \log p + \beta \|\tilde{\lambda}\|_{TV} \\ &\leq \left( \sum_{k=0}^{K-1} \sum_{i=1}^{\hat{D}} C_{k,i}(\tilde{\lambda}) - \tilde{y}_{k,i}^E \log C_{k,i}^E(\tilde{\lambda}) - \tilde{y}_{k,i}^S \log(p \cdot C_{k,i}^S(\tilde{\lambda})) \right) + b \log p + \beta \|\tilde{\lambda}\|_{TV} \\ &= \sum_{k=0}^{K-1} \sum_{i=1}^{\hat{D}} C_{k,i}(\tilde{\lambda}) - \tilde{y}_{k,i}^E \log C_{k,i}^E(\tilde{\lambda}) - \tilde{y}_{k,i}^S \log C_{k,i}^S(\tilde{\lambda}) + \beta \|\tilde{\lambda}\|_{TV} \\ &= \tilde{J}^{\tilde{y}^E, \tilde{y}^S}(\tilde{\lambda}) . \end{aligned}$$

Über mehrere Schritte haben wir somit eine Ungleichungskette für die Minimierungsfunktionale erhalten:

$$\tilde{J}^{\tilde{y}^E, \tilde{y}^S}(\tilde{\lambda}) \leq \tilde{J}^{\tilde{y}^E, \tilde{y}^S}(\hat{\lambda}) \leq \hat{J}^{y, p}(\hat{\lambda}) + b \log p \leq \hat{J}^{y, p}(\tilde{\lambda}) + b \log p \leq \tilde{J}^{\tilde{y}^E, \tilde{y}^S}(\tilde{\lambda}).$$

Es folgt die Gleichheit zwischen den Termen und  $\hat{\lambda}$  ist für  $\tilde{J}^{\tilde{y}^E, \tilde{y}^S}$  ebenfalls ein Minimierer. Andersherum ist  $\tilde{\lambda}$  auch ein Minimierer von  $\hat{J}^{y, p}$ . Insgesamt beweist dies die Gleichheit zwischen den Mengen aller Minimierer von  $\tilde{J}^{\tilde{y}^E, \tilde{y}^S}$  und  $\hat{J}^{y, p}$ :

$$\left\{ \lambda \in M(\Omega) \mid \lambda = \underset{\mu \in M(\Omega)}{\operatorname{argmin}} \tilde{J}^{\tilde{y}^E, \tilde{y}^S}(\mu) \right\} = \left\{ \lambda \in M(\Omega) \mid \lambda = \underset{\mu \in M(\Omega)}{\operatorname{argmin}} \hat{J}^{y, p}(\mu) \right\}.$$

Wir dürfen daher für die Suche nach Minimierern des numerisch schwer zu lösenden Problems  $\tilde{J}^{\tilde{y}^E, \tilde{y}^S}$  das gewichtete Maximums-Funktional  $\hat{J}^{y, p}$  benutzen. Es bleibt eine letzte Optimierung vorzunehmen.

## 4.4 Konvex relaxiertes Rekonstruktionsfunktional

Durch den Term  $-\log \left( \max(p \cdot C_{k,i}^S(\cdot), C_{k,i}^E(\cdot)) \right)$  ist  $\hat{J}^{y, p}$  weder ein konvexes Funktional noch besitzt es die Eigenschaft der Unterhalbstetigkeit bezüglich einer (möglicherweise schwächeren) Topologie. Allerdings ist die Konvexität des Rekonstruktionsfunktionals wünschenswert, um effizientere numerische Lösungsverfahren nutzen zu können. Des Weiteren wird für die Anwendung der direkten Methode die Unterhalbstetigkeit bezüglich der schwach- $*$ -Topologie benötigt. Wie in Abschnitt 2.2 erwähnt, greifen wir auf eine relaxierte Funktion zurück: Dabei ersetzen wir  $\max(\dots)$  durch die konvexe Hülle des Maximums, wonach die nötige Bedingung 2.4 gilt. Die schwach- $*$ -Unterhalbstetigkeit, ebenso wie die Koerzivität, beweisen wir später. Letztendlich erhalten wir das finale Rekonstruktionsfunktional

$$J^{y, p} : M(\Omega) \longrightarrow \mathbb{R}$$

mit der Funktionsvorschrift

$$\lambda \longmapsto J^{y, p}(\lambda) := \sum_{k=0}^{K-1} \sum_{i=1}^{\hat{D}} C_{k,i}(\lambda) - y_{k,i} \log(p \cdot C_{k,i}^S(\lambda) + C_{k,i}^E(\lambda)) + \beta \|\lambda\|_{TV}, \quad (4.9)$$

wobei die Messungen  $y = (y_{k,i})_{k,i}$  Realisierungen von  $Y$  sind,  $p \geq 0$  der Streuungsparameter ist und  $\beta > 0$  gilt. Interessanterweise spielt die Anzahl  $b \in \mathbb{N}$  an Messereignissen, die durch gestreute Photonen ereignet sind, keine Rolle im Minimierungsfunktional, ebenso wie die genaue Aufteilung der Messungen in gestreut und perfekt fliegend. Die Unterscheidung dazwischen wird durch eine geeignete Wahl des Streuungsparameters  $p$  vorgenommen und letzten Endes entscheidet jene Wahl über die erfolgreiche Rekonstruktion des Maßes 4.1,  $\hat{\lambda} \in M(\Omega)$ .

Zusammenfassend haben wir durch

$$\min_{\lambda \in M(\Omega)} J^{y, p}(\lambda) \quad (4.10)$$

ein Minimierungsproblem aufgestellt, das die Schlechtgestellttheit des zugrunde liegenden Problems löst, die Positionen und Geschwindigkeiten aller Isoptope zu rekonstruieren. Mit

$$\lambda^* \in \operatorname{argmin}_{\lambda \in M(\Omega)} J^{y,p}(\lambda)$$

erhalten wir schlussendlich die Repräsentation des Maßes der Radionuklide auf  $[0, 1]^d$  und besagte Startpositionen  $x_i$  und Geschwindigkeiten  $v_i$  eines jeden Nuklides lassen sich ablesen.



## 5 Existenz von Minimierern

Bisher sind wir stets von der Existenz eines Minimierers  $\lambda^*$  für die verschiedenen in Kapitel 4 hergeleiteten Funktionen ausgegangen. Diese ist jedoch nicht trivialerweise gegeben. Mittels der in Abschnitt 2.2 vorgestellten direkten Methode haben wir uns einen Ansatz erarbeitet, der uns die Existenz und Eindeutigkeit eines Minimierers garantiert. In diesem Kapitel werden die nötigen Voraussetzungen für die Nutzung des zentralen Satzes 2.19 bewiesen, nämlich:

1. Der Phasenraum ist normiert, lokalkompakt und ein Hausdorff-Raum.
2.  $J^{y,p}$  ist wohldefiniert.
3. Es existiert eine untere Schranke für  $J^{y,p}$ .
4.  $J^{y,p}$  ist koerziv.
5.  $J^{y,p}$  ist schwach- $*$ -unterhalbstetig.
6.  $J^{y,p}$  ist (strikt) konvex.

Sind diese Bedingungen erfüllt, so existiert nach der direkten Methode ein eindeutiges Maß  $\lambda^* \in M(\Omega)$ , das der Gleichung  $J^{y,p}(\lambda^*) = \min_{\lambda \in M(\Omega)} J^{y,p}(\lambda)$  genügt. Wir werden jedoch sehen, dass mit  $\lambda^* \in M(\Omega)$  lediglich eine konvexe Funktion vorliegt, damit einhergehend die Eindeutigkeit nicht gewährleistet ist. Im Folgenden analysieren wir die einzelnen Punkte der Reihenfolge nach und beginnen mit dem Phasenraum.

1. Als Teilraum des  $\mathbb{R}^d \times \mathbb{R}^d$  ist die Normiertheit und Hausdorff-Eigenschaft von  $\Omega$  gegeben. Für die Lokalkompaktheit zeigen wir die Kompaktheit. Wir können den Phasenraum leicht umformulieren, indem wir die möglichen Geschwindigkeitsvektoren genauer einschränken. Ein am Rand startendes Element  $x = 0 \in \mathbb{R}^d$  kann sich linear maximal mit der Geschwindigkeit  $\frac{1}{T} \in \mathbb{R}$  in jede (positive) Richtung der  $d$ -Dimensionen bewegen. Andernfalls würde für einen Geschwindigkeitsvektor, der mindestens eine Komponente  $v_j > \frac{1}{T}$  besitzt,  $j \in \{1, \dots, d\}$ , folgende Implikation gelten:

$$x + T \sum_{j=1}^d v_j \cdot e_j = T \sum_{j=1}^d v_j \cdot e_j \notin [0, 1]^d .$$

Mit  $e_j$  bezeichnen wir hier den  $j$ -ten Einheitsvektor. Dem gleichen Gedankengang nach ist auch die Geschwindigkeit von unten begrenzt und es gilt:

$$\begin{aligned}\Omega &= \left\{ (x, v) \in [0, 1]^d \times \mathbb{R}^d \mid x + tv \in [0, 1]^d, \text{ für alle } t \in [0, T] \right\} \\ &= \left\{ (x, v) \in [0, 1]^d \times \left[ \frac{-1}{T}, \frac{1}{T} \right]^d \mid x + tv \in [0, 1]^d, \text{ für alle } t \in [0, T] \right\} .\end{aligned}$$

Der Phasenraum ist daher ein beschränkter Raum. Für die Abgeschlossenheit betrachten wir eine Folge  $(x_n, v_n)_n \subset \Omega$  mit  $\lim_{n \rightarrow \infty} (x_n, v_n)_n = (x, v) \in \mathbb{R}^d \times \mathbb{R}^d$ . Die Konvergenz ist komponentenweise zu verstehen und hierüber erhalten wir auch die gewünschte Aussage. Da  $[0, 1] \subset \mathbb{R}$  und  $\left[ \frac{-1}{T}, \frac{1}{T} \right] \subset \mathbb{R}$  abgeschlossen sind, sind die komponentenweisen Grenzwerte ebenfalls Element aus  $[0, 1]$ , bzw.  $\left[ \frac{-1}{T}, \frac{1}{T} \right]$ . Weiter gilt für alle  $n \in \mathbb{N}$  und  $t \in [0, T]$ , dass  $x_n + tv_n \in [0, 1]^d$  und  $(x_n + tv_n)_n$  ist eine Folge in  $[0, 1]^d$ . Aufgrund der Abgeschlossenheit von  $[0, 1]^d$  folgt erneut die Existenz des Grenzwertes in  $[0, 1]^d$ . Insgesamt gilt:

$$\begin{aligned}\lim_{n \rightarrow \infty} (x_n + tv_n) &= \lim_{n \rightarrow \infty} x_n + t \lim_{n \rightarrow \infty} v_n = x + tv \in [0, 1]^d, \\ \text{und } \lim_{n \rightarrow \infty} x_n &= x \in [0, 1]^d, \quad \lim_{n \rightarrow \infty} v_n = v \in \left[ \frac{-1}{T}, \frac{1}{T} \right]^d.\end{aligned}$$

Somit ist  $(x, v)$  als Grenzwert ein Element des Phasenraumes und  $\Omega$  abgeschlossen. Nach dem Satz von Heine-Borel folgt die Kompaktheit von  $\Omega$ .

2. Um die Wohldefiniertheit von

$$J^{y,p}(\lambda) = \sum_{k=0}^{K-1} \sum_{i=1}^{\widehat{D}} C_{k,i}(\lambda) - y_{k,i} \log(p \cdot C_{k,i}^S(\lambda) + C_{k,i}^E(\lambda)) + \beta \|\lambda\|_{TV}$$

zu zeigen, betrachten wir zunächst  $C_{k,i}^S$  und  $C_{k,i}^E$  und zeigen deren Wohldefiniertheit. Für  $C_{k,i}^S$  gilt mit dem Transformationssatz 2.1:

$$\begin{aligned}C_{k,i}^S(\lambda) &= dt \times p_S F^S((\phi_t)_\# \lambda)([t_k, t_{k+1}] \times \Gamma_{(i)}) = \int_{t_k}^{t_{k+1}} p_S F^S((\phi_t)_\# \lambda)(\Gamma_{(i)}) dt \\ &= p_S \int_{t_k}^{t_{k+1}} (\phi_t)_\# \lambda([0, 1]^d) \cdot \underbrace{\frac{(H^{d-1} \times H^{d-1})_{(\partial[0,1]^d \times \partial[0,1]^d)}(\Gamma_{(i)})}{(H_{\partial[0,1]^d}^{d-1}(\partial[0,1]^d))^2}}_{\leq 1} dt \\ &\leq p_S \int_{t_k}^{t_{k+1}} \int_{[0,1]^d} 1 d(\phi_t)_\# \lambda dt \stackrel{2.1}{=} p_S \int_{t_k}^{t_{k+1}} \int_{\Omega} \phi_t d\lambda dt \\ &\leq \frac{p_S}{K} \|\phi_t\|_{\infty} \|\lambda\|_{TV} .\end{aligned}$$

Genauer gilt für die Supremumsnorm von  $\phi_t$ :

$$\sup_{(x,v) \in \Omega} \|\phi_t((x, v))\|_1 = \sup_{(x,v) \in \Omega} \|x + tv\|_1 = \sup_{(x,v) \in \Omega} \sum_{i=1}^d \underbrace{|x_i + tv_i|}_{\in [0,1]} = d .$$

Wir haben hier die Summennorm auf  $\mathbb{R}^d$  betrachtet und erhalten insgesamt durch  $C_{k,i}^S(\lambda) \leq \frac{d \cdot p_S}{K} \|\lambda\|_{TV} < \infty$  die Wohldefiniertheit von  $C_{k,i}^S$ .

Ähnliche Schritte führen zu der Wohldefiniertheit von  $C_{k,i}^E$ . Doch vor der Rechnung benötigen wir einen genaueren Blick auf die Auswertung des Erfassungoperator  $F^E(\mu)$ , wie er in Abschnitt 3.2 vorgestellt wurde. Hierfür ist  $\mu \in M([0, 1]^d)$ . Dann gilt

$$\begin{aligned} F^E(\mu)(\Gamma_{(i)}) &= F_3^E F_2^E F_1^E(\mu)(\Gamma_{(i)}) = P_{\#}(F_2^E F_1^E(\mu))(\Gamma_{(i)}) = F_2^E F_1^E(\mu)(P^{-1}(\Gamma_{(i)})) \\ &= (F_1^E(\mu) \times v_d)(P^{-1}(\Gamma_{(i)})) = \left( \int_{[0,1]^d} G_x(y) d\mu(x) \times v_d \right) (P^{-1}(\Gamma_{(i)})) . \end{aligned}$$

An dieser Stelle betrachten rufen wir uns die Struktur von  $P^{-1}(\Gamma_{(i)})$  in Erinnerung (Abschnitt 3.2):

$$P^{-1}(\Gamma_{(i)}) = \{(z, v) \in [0, 1]^d \times G(1, d) \mid P((z, v)) \in \Gamma_{(i)}\} .$$

Für das weitere Ausschreiben des Operators benötigen wir die Definition einer Menge, die für eine feste Richtungskomponente  $v \in G(1, d)$  alle zugehörigen Positionen  $x \in [0, 1]^d$  erfasst. Dementsprechend sei

$$P_z^{-1}(\Gamma_{(i)}, v) := \{z \in [0, 1]^d \mid (z, v) \in P^{-1}(\Gamma_{(i)})\} ,$$

und nach dem Prinzip von Cavalieri ist  $P_z^{-1}(\Gamma_{(i)}, v)$  messbar für alle  $v \in G(1, d)$ . Durch das selbe Prinzip ergibt sich auch die logische Aufschlüsselung von  $F^E$  in seine Bestandteile als

$$\begin{aligned} F^E(\mu)(\Gamma_{(i)}) &= \left( \int_{[0,1]^d} G_x(y) d\mu(x) \times v_d \right) (P^{-1}(\Gamma_{(i)})) \\ &= \int_{G(1,d)} \int_{P_z^{-1}(\Gamma_{(i)},v)} \int_{[0,1]^d} G_x(y) d\mu(x) dy dv_d . \end{aligned}$$

Nun können wir mit dem Beweis der Wohldefiniertheit fortfahren:

$$\begin{aligned} C_{k,i}^E(\lambda) &= dt \times p_E F^E((\phi_t)_{\#}\lambda)([t_k, t_{k+1}] \times \Gamma_{(i)}) \\ &= p_E \int_{t_k}^{t_{k+1}} \int_{G(1,d)} \int_{P_z^{-1}(\Gamma_{(i)},v)} \int_{[0,1]^d} G_x(y) d(\phi_t)_{\#}\lambda(x) dy dv_d dt \\ &\leq p_E \int_{t_k}^{t_{k+1}} \int_{G(1,d)} \int_{[0,1]^d} \int_{[0,1]^d} G_x(y) d(\phi_t)_{\#}\lambda(x) dy dv_d dt \\ &\stackrel{2.1}{=} p_E \int_{G(1,d)} \int_{[0,1]^d} \int_{\Omega} G_x(y) \circ \phi_t(x) d\lambda(x) dy dv_d dt \\ &\stackrel{\text{Fubini}}{=} p_E \int_{t_k}^{t_{k+1}} \int_{G(1,d)} \underbrace{\int_{\Omega} \int_{[0,1]^d} G_{\phi_t(x)}(y) dy d\lambda(x)}_{= 1 \text{ für alle } x \in \Omega} dv_d dt \\ &= p_E \int_{t_k}^{t_{k+1}} dt \int_{\Omega} d\lambda \underbrace{\int_{G(1,d)} dv_d}_{= 1 \text{ nach Definition}} \\ &\leq \frac{p_E}{K} \|\lambda\|_{TV} < \infty . \end{aligned}$$

Wir haben hier benutzt, dass  $v_d$  als Haar-Maß ein Wahrscheinlichkeitsmaß ist. Insgesamt folgt  $C_{k,i}(\lambda) = C_{k,i}^E(\lambda) + C_{k,i}^S(\lambda) < \infty$  und somit ist neben der Wohldefiniertheit auch die Stetigkeit von  $C_{k,i}$  bewiesen. Diese folgt aus der Linearität von  $F$  als Vorwärtsoperator und der eben gezeigten Beschränktheit. Um mit dem Beweis der Wohldefiniertheit von  $J^{y,p}$  fortzufahren, benötigen wir noch zwei Resultate.

Zum einen sei angemerkt, dass für alle  $\lambda \in M(\Omega)$  ungleich dem Null-Maß auf  $M(\Omega)$  gilt:

$$C_{k,i}(\lambda) \geq C_{k,i}^S(\lambda) = p_S \int_{t_k}^{t_{k+1}} \underbrace{(\phi_t)_\# \lambda([0, 1]^d)}_{> 0} \cdot \underbrace{\frac{(H^{d-1} \times H^{d-1})_{(\partial[0,1]^d \times \partial[0,1]^d)}(\Gamma(i))}{(H_{\partial[0,1]^d}^{d-1}(\partial[0, 1]^d))^2}}_{> 0} dt > 0 ,$$

für alle  $k$  und  $i$ . Für den Fall  $\lambda = 0$  setzen wir  $J(0) := +\infty$ . Somit fällt dem Null-Maß bei der Minimierung keine Rolle zu. Auch von der Anwendungsseite ist das Null-Maß nicht zu beachten, schließlich würde dieses zu keinen Messungen führen.

Zum anderen betrachten wir die möglichen Werte der Messungen  $y_{k,i}$ . Als Realisierung einer  $\text{Poi}(C_{k,i}(\lambda))$ -verteilten Zufallsvariable  $Y_{k,i}$  können wir die Wahrscheinlichkeit berechnen, dass ausschließlich endliche Werte ermittelt werden. Hierfür nutzen wir die Darstellung der Exponentialfunktion durch die Exponentialreihe.

$$\begin{aligned} \mathbb{P}[Y_{k,i} < \infty] &= \mathbb{P}\left[\bigcup_{n=0}^{\infty} \{Y_{k,i} = n\}\right] = \sum_{n=0}^{\infty} \mathbb{P}[Y_{k,i} = n] \\ &= \sum_{n=0}^{\infty} \frac{1}{n!} C_{k,i}(\lambda)^n e^{-C_{k,i}(\lambda)} = e^{-C_{k,i}(\lambda)} \cdot e^{C_{k,i}(\lambda)} \\ &= 1 . \end{aligned} \tag{5.1}$$

Somit ist  $y_{k,i} < \infty$  fast sicher. Insgesamt folgt für unser betrachtetes Minimierungsfunktional die Wohldefiniertheit:

$$\begin{aligned} J^{y,p}(\lambda) &= \sum_{k=0}^{K-1} \sum_{i=1}^{\hat{D}} C_{k,i}(\lambda) - y_{k,i} \log(p \cdot C_{k,i}^S(\lambda) + C_{k,i}^E(\lambda)) + \beta \|\lambda\|_{TV} \\ &\leq \sum_{k=0}^{K-1} \sum_{i=1}^{\hat{D}} \underbrace{C_{k,i}(\lambda)}_{< \infty} - \underbrace{y_{k,i}}_{< \infty} \underbrace{\log(p \cdot C_{k,i}^S(\lambda))}_{\in (0, \infty)} + \beta \|\lambda\|_{TV} < \infty . \end{aligned}$$

3. Als nächsten Schritt gilt es eine untere Schranke herzuleiten. Hierfür betrachten wir eine kurze Fallunterscheidung. Für  $p \geq 1$  folgt

$$\begin{aligned} J^{y,p}(\lambda) &\geq \sum_{k=0}^{K-1} \sum_{i=1}^{\hat{D}} C_{k,i}(\lambda) - y_{k,i} \log(p \cdot C_{k,i}^S(\lambda) + C_{k,i}^E(\lambda)) \\ &\geq \sum_{k=0}^{K-1} \sum_{i=1}^{\hat{D}} C_{k,i}(\lambda) - y_{k,i} \log(p \cdot C_{k,i}(\lambda)) \end{aligned}$$

$$= \sum_{k=0}^{K-1} \sum_{i=1}^{\widehat{D}} C_{k,i}(\lambda) - y_{k,i} \log p - y_{k,i} \log (C_{k,i}(\lambda))$$

und andernfalls, für  $p < 1$ :

$$J^{y,p}(\lambda) \geq \sum_{k=0}^{K-1} \sum_{i=1}^{\widehat{D}} C_{k,i}(\lambda) - y_{k,i} \log (C_{k,i}(\lambda)) .$$

Durch die Beschränktheit von unten von  $J^{y,p}$  für  $p < 1$  ist auch die Beschränktheit von unten für  $p \geq 1$  gegeben, wobei die Summe über die Terme  $y_{k,i} \log p < \infty$  die untere Schranke verschiebt. Für die weitere Abschätzung konstruieren wir einen maximalen Messwert für alle Zeitpunkte und für jedes Detektorenpaar, und definieren

$$y_{\max} := \max \left( 1, \max_{k,i} y_{k,i} \right) .$$

Nach Gleichung 5.1 ist  $y_{\max}$  fast sicher endlich. Wir ersetzen die Messungen  $y_{k,i}$  im nächsten Schritt und erhalten

$$J^{y,p}(\lambda) \geq \sum_{k=0}^{K-1} \sum_{i=1}^{\widehat{D}} C_{k,i}(\lambda) - y_{\max} \log (C_{k,i}(\lambda)) .$$

Da  $C_{k,i}(\cdot) \in (0, \infty)$  gilt, sind die Summanden von der Form einer Funktion

$$\begin{aligned} f : \mathbb{R}_+ &\longrightarrow \mathbb{R} \\ x &\longmapsto x - y_{\max} \log x . \end{aligned}$$

Eine elementare Kurvendiskussion zeigt, dass  $f$  ein globales Minimum an der Stelle  $x = y_{\max}$  annimmt. Dieses Wissen angewandt auf die Summanden erhalten wir

$$\begin{aligned} J^{y,p}(\lambda) &\geq \sum_{k=0}^{K-1} \sum_{i=1}^{\widehat{D}} C_{k,i}(\lambda) - y_{\max} \log (C_{k,i}(\lambda)) \\ &\geq \sum_{k=0}^{K-1} \sum_{i=1}^{\widehat{D}} y_{\max} - y_{\max} \log (y_{\max}) = \frac{K \cdot D(D+1)}{2} (y_{\max} (1 - \log (y_{\max}))) =: \tilde{c} . \end{aligned}$$

Das Minimierungsfunktional ist durch  $\tilde{c}$  von unten beschränkt.

4. Durch die Existenz einer unteren Schranke werden nur wenige Schritte benötigt, um die Koerzivität zu zeigen. Sei  $(\mu_n)_n \subset M(\Omega)$  eine Folge von Maßen mit  $\|\mu_n\|_{TV} \xrightarrow{n \rightarrow \infty} \infty$ . Mit der Abschätzung nach unten mittels  $\tilde{c}$  erhalten wir

$$J^{y,p}(\lambda) \geq \tilde{c} + \beta \|\mu_n\|_{TV} \xrightarrow{n \rightarrow \infty} \infty .$$

Demnach ist das Funktional  $J^{y,p}$  koerzitiv.

5. Ähnlich zur Wohldefiniertheit betrachten wir erneut zunächst  $C_{k,i}^E$  und  $C_{k,i}^S$ . Dabei werden wir deren schwach-\*-Stetigkeit beweisen, die anschließend schnell hin zur schwach-\*-Unterhalbstetigkeit von  $J^{y,p}$  führt. Wir beginnen mit einer Folge schwach-\*-konvergenter Folge  $(\lambda_n)_n \subset M(\Omega)$ ,  $\lambda_n \xrightarrow{w^*} \lambda \in M(\Omega)$ . Dies bedeutet

$$\int_{\Omega} f d\lambda_n \xrightarrow{n \rightarrow \infty} \int_{\Omega} f d\lambda \quad \text{für alle } f \in C_0(\Omega) ,$$

und somit insbesondere

$$\lambda_n(\Omega) = \int_{\Omega} 1 d\lambda_n \xrightarrow{n \rightarrow \infty} \int_{\Omega} 1 d\lambda = \lambda(\Omega) .$$

Da  $\Omega$  eine kompakte Menge ist, gilt  $C_0(\Omega) = C_b(\Omega) = C_c(\Omega)$ , und für alle  $k$  ist auch  $\phi_t$  als beschränkte (stetige und lineare) Funktion auf  $\Omega$  Element aus  $C_b(\Omega)$ . Weiter folgt damit die Konvergenz der Massen des Bildmaßes von  $\lambda_n$  unter  $\phi_t$  gegen die des Bildmaßes von  $\lambda$  unter  $\phi_t$ :

$$\begin{aligned} (\phi_t)_{\#} \lambda_n([0, 1]^d) &= \int_{[0,1]^d} 1 d(\phi_t)_{\#} \lambda_n \stackrel{2.1}{=} \int_{\Omega} \phi_t d\lambda_n \\ &\xrightarrow{n \rightarrow \infty} \int_{\Omega} \phi_t d\lambda \stackrel{2.1}{=} (\phi_t)_{\#} \lambda([0, 1]^d) . \end{aligned}$$

Mit dieser Gleichung erhalten wir folgende Rechnung für den gestreuten Vorwärtsoperator:

$$\begin{aligned} C_{k,i}^S(\lambda_n) &= p_S \int_{t_k}^{t_{k+1}} (\phi_t)_{\#} \lambda_n([0, 1]^d) \cdot \frac{(H^{d-1} \times H^{d-1})_{(\partial[0,1]^d \times \partial[0,1]^d)}(\Gamma(i))}{(H_{\partial[0,1]^d}^{d-1}(\partial[0, 1]^d))^2} dt \\ &\xrightarrow{n \rightarrow \infty} p_S \int_{t_k}^{t_{k+1}} (\phi_t)_{\#} \lambda([0, 1]^d) \cdot \frac{(H^{d-1} \times H^{d-1})_{(\partial[0,1]^d \times \partial[0,1]^d)}(\Gamma(i))}{(H_{\partial[0,1]^d}^{d-1}(\partial[0, 1]^d))^2} dt \\ &= C_{k,i}^S(\lambda) . \end{aligned}$$

Dies beweist die schwach-\*-Stetigkeit von  $C_{k,i}^S$ . Im nächsten Schritt wenden wir uns  $C_{k,i}^E$  zu. Dabei betrachten wir den Gauß-Filter  $G_x(y)$  für ein festes  $y \in [0, 1]^d$ , mit der Verkettung von  $\phi_t \in C_b(\Omega)$  und  $\epsilon > 0$ :

$$\begin{aligned} G_{(\cdot)}(y) \circ \phi_t(\cdot) &= G_{\phi_t(\cdot)}(y) \\ &= \frac{1}{(\sqrt{2\pi}\epsilon)^2} \cdot \exp\left(-\frac{\|y - \phi_t(\cdot)\|_2^2}{2\epsilon^2}\right) \in C_b(\Omega) . \end{aligned}$$

Die Beschränktheit und Stetigkeit sind offensichtlich und es folgt

$$\begin{aligned} \int_{[0,1]^d} G_x(y) d(\phi_t)_{\#} \lambda_n &\stackrel{2.1}{=} \int_{\Omega} G_{\phi_t(x)}(y) d\lambda_n \\ &\xrightarrow{n \rightarrow \infty} \int_{\Omega} G_{\phi_t(x)}(y) d\lambda = \int_{[0,1]^d} G_x(y) d(\phi_t)_{\#} \lambda . \end{aligned}$$

Das Resultat eingesetzt in  $C_{k,i}^E$  führt uns zum gewünschten Ergebnis, nämlich

$$\begin{aligned} C_{k,i}^E(\lambda_n) &= p_E \int_{t_k}^{t_{k+1}} \int_{G(1,d)} \int_{P_z^{-1}(\Gamma_{(i),v})} \int_{[0,1]^d} G_x(y) d(\phi_t)_\# \lambda_n(x) dy dv_d dt \\ &\xrightarrow{n \rightarrow \infty} p_E \int_{t_k}^{t_{k+1}} \int_{G(1,d)} \int_{P_z^{-1}(\Gamma_{(i),v})} \int_{[0,1]^d} G_x(y) d(\phi_t)_\# \lambda(x) dy dv_d dt = C_{k,i}^E(\lambda) . \end{aligned}$$

Mit Satz 2.18 und den zuvor gemachten Aussagen über die Komponenten von  $C_{k,i}$  gilt für eine schwach- $*$ -konvergente Folge  $(\lambda_n)_n \subset M(\Omega)$

$$\begin{aligned} \liminf_{n \rightarrow \infty} J^{y,p}(\lambda_n) &= \liminf_{n \rightarrow \infty} \left( \sum_{k=0}^{K-1} \sum_{i=1}^{\widehat{D}} C_{k,i}(\lambda_n) - y_{k,i} \log (p \cdot C_{k,i}^S(\lambda_n) + C_{k,i}^E(\lambda_n)) \right. \\ &\quad \left. + \beta \|\lambda_n\|_{TV} \right) \\ &\geq \sum_{k=0}^{K-1} \sum_{i=1}^{\widehat{D}} \liminf_{n \rightarrow \infty} \left( C_{k,i}(\lambda_n) - y_{k,i} \log (p \cdot C_{k,i}^S(\lambda_n) + C_{k,i}^E(\lambda_n)) \right) \\ &\quad + \liminf_{n \rightarrow \infty} (\beta \|\lambda_n\|_{TV}) \\ &\geq \sum_{k=0}^{K-1} \sum_{i=1}^{\widehat{D}} \liminf_{n \rightarrow \infty} C_{k,i}(\lambda_n) - y_{k,i} \log \limsup_{n \rightarrow \infty} (p \cdot C_{k,i}^S(\lambda_n) + C_{k,i}^E(\lambda_n)) \\ &\quad + \beta \|\lambda\|_{TV} \\ &= \sum_{k=0}^{K-1} \sum_{i=1}^{\widehat{D}} C_{k,i}(\lambda) - y_{k,i} \log (p \cdot C_{k,i}^S(\lambda) + C_{k,i}^E(\lambda)) + \beta \|\lambda\|_{TV} \\ &= J^{y,p}(\lambda) . \end{aligned}$$

Damit ist auch die schwach- $*$ -Unterhalbstetigkeit des Minimierungsfunktionals gezeigt. An dieser Stelle können wir bereits die Existenz eines Minimierers folgern.

6. Für die (strikte) Konvexität betrachten wir zwei Maße  $\mu, \lambda \in M(\Omega)$  und deren Konvexkombination  $b \cdot \mu + \tilde{b} \cdot \lambda$ , die selbstredend erneut Element aus  $M(\Omega)$  ist, wobei  $\tilde{b} = 1 - b$  und  $b \in [0, 1]$  gilt. Wir greifen in der Abschätzung auf die Linearität von  $C_{k,i}^E$  und  $C_{k,i}^S$ , ferner

auf die Dreiecksungleichung der TV-Norm zurück und erhalten

$$\begin{aligned}
 J^{y,p}(b \cdot \mu + \tilde{b} \cdot \lambda) &= \sum_{k=0}^{K-1} \sum_{i=1}^{\hat{D}} C_{k,i} \left( b \cdot \mu + \tilde{b} \cdot \lambda \right) \\
 &\quad - y_{k,i} \log \left( p \cdot C_{k,i}^S(b \cdot \mu + \tilde{b} \cdot \lambda) + C_{k,i}^E(b \cdot \mu + \tilde{b} \cdot \lambda) \right) \\
 &\quad + \beta \|b \cdot \mu + \tilde{b} \cdot \lambda\|_{TV} \\
 &\leq \sum_{k=0}^{K-1} \sum_{i=1}^{\hat{D}} b \cdot C_{k,i}(\mu) + \tilde{b} C_{k,i}(\lambda) \\
 &\quad - y_{k,i} \log \left( p \cdot b \cdot C_{k,i}^S(\mu) + p \cdot \tilde{b} \cdot C_{k,i}^S(\lambda) + b \cdot C_{k,i}^E(\mu) + \tilde{b} \cdot C_{k,i}^E(\lambda) \right) \\
 &\quad + b \cdot \beta \|\mu\|_{TV} + \tilde{b} \cdot \beta \|\lambda\|_{TV} .
 \end{aligned}$$

An dieser Stelle wollen wir den Logarithmus abschätzen. Auf dem Intervall  $(0, \infty)$  ist der negative Logarithmus strikt konvex, jedoch gilt für eine lineare, stetige und injektive Funktion  $f$  mit  $f(x) = f(y)$

$$f(b \cdot x + \tilde{b} \cdot y) = b \cdot f(x) + \tilde{b} \cdot f(y) = f(x) = f(y) ,$$

woraus wir die Gleichungskette

$$\begin{aligned}
 \log \left( f(b \cdot x + \tilde{b} \cdot y) \right) &= \log (f(y)) = \log (f(x)) \\
 &= (b + \tilde{b}) \cdot \log f(x) = b \cdot \log (f(x)) + \tilde{b} \cdot \log (f(y))
 \end{aligned}$$

erhalten, nach der keine strikte Konkavität für den Logarithmus angewandt auf  $f$  gilt. Da die Komposition  $p \cdot C_{k,i}^S + C_{k,i}^E$  ebenfalls injektiv ist, können wir daher lediglich die Konvexität beweisen:

$$\begin{aligned}
 J^{y,p}(b \cdot \mu + \tilde{b} \cdot \lambda) &\leq b \cdot \left( \sum_{k=0}^{K-1} \sum_{i=1}^{\hat{D}} C_{k,i}(\mu) - y_{k,i} \log (p \cdot C_{k,i}^S(\mu) + C_{k,i}^E(\mu)) + \beta \|\mu\|_{TV} \right) \\
 &\quad + \tilde{b} \cdot \left( \sum_{k=0}^{K-1} \sum_{i=1}^{\hat{D}} C_{k,i}(\lambda) - y_{k,i} \log (p \cdot C_{k,i}^S(\lambda) + C_{k,i}^E(\lambda)) + \beta \|\lambda\|_{TV} \right) \\
 &= b \cdot J^{y,p}(\mu) + \tilde{b} \cdot J^{y,p}(\lambda) .
 \end{aligned}$$

Dass  $p \cdot C_{k,i}^S + C_{k,i}^E$  im Allgemeinen injektiv ist, kann durch einfach gehaltene Konstruktionen gezeigt werden. Zum Beispiel mit der Wahl von zwei Detektoren, wonach  $\hat{D} = 1$  gilt und einer Messung,  $K = 1$ . Des Weiteren seien zwei verschiedene Positionen eines Nuklids gegeben, repräsentiert durch  $\mu, \lambda \in M(\Omega)$ . Die Positionen seien so gewählt, dass das Radionuklid sich in einer jeweils anderen Ecke des Tomographen  $[0, 1]^d$  befindet. Da nur eine Messung vorgenommen wird, kommt der Bewegungsrichtung keine Bedeutung zu und für den

Streuungsterm folgt trivialerweise  $C_{k,i}^S(\mu) = C_{k,i}^S(\lambda)$ . Ebenfalls gilt die Gleichheit zwischen den Erfassungstermen  $C_{k,i}^E(\mu) = C_{k,i}^E(\lambda)$  und der Term  $p \cdot C_{k,i}^S + C_{k,i}^E$  ist injektiv.

Nach erstens bis fünftens sind die Voraussetzungen für die Existenz eines Minimierers nach Satz 2.19 erfüllt. Die strikte Konvexität ist nicht gegeben, wonach wir nicht auf die Eindeutigkeit dieser Minimierer schließen können. Dennoch ist es für das folgende Kapitel hilfreich, dass die Konvexität von  $J^{y,p}$  bei sechstens gezeigt wurde.



## 6 Numerische Simulation

Aufgrund der Komplexität der Funktion  $J^{y,p}$  ist die Ermittlung ihres Minimierers rechnerisch per Hand kaum zu bewältigen. Für einen Computer ist dies im Gegensatz dazu möglich, zumal das Minimierungsproblem 4.10

$$\min_{\lambda \in M(\Omega)} J^{y,p}(\lambda)$$

ein konvexes Optimierungsproblem beschreibt. Ein gängiger Algorithmus zur Lösung solcher Probleme ist der Frank-Wolfe-Algorithmus, doch bevor wir diesen anwenden können benötigen wir eine Diskretisierung des Minimierungsproblems 4.10. Das Ziel dieses Kapitels besteht darin, die nötigen Umformungen vorzustellen und abschließend in zwei Dimensionen eine beispielhafte Simulation des Frank-Wolfe-Algorithmus zu implementieren. Als Quellen dienen hauptsächlich „Revisiting Frank-Wolfe: Projection-Free Sparse Convex Optimization“ [12] von Martin Jaggi, und „An algorithm for quadratic programming“ [17] von Philip Wolfe und Marguerite Frank. Weiter sei auf das Buch „Numerical Optimization“ [13] von Jorge Nocedal und Stephen J. Wright verwiesen.

### 6.1 Frank-Wolfe-Algorithmus

Wir betrachten ein konvexes Minimierungsproblem der Form

$$\min_{x \in D} f(x) .$$

Hierbei ist  $f : D \rightarrow \mathbb{R}$  eine konvexe und differenzierbare Funktion, ferner  $D$  eine kompakte und konvexe Teilmenge des  $\mathbb{R}^d$ . In der 1956 erschienenen Arbeit [17] beschreiben Frank und Wolfe die nach ihnen benannte iterative Methode für die Lösung jener Optimierungsprobleme, auch bekannt als *conditional gradient method*.

Das Verfahren beginnt mit einem beliebigen Startpunkt  $x_0 \in D$  und durch Iterationen nähert es sich dem Minimum von  $f$ , indem es im  $k$ -ten Schritt mit Hilfe des zuvor berechneten  $x_{k-1} \in D$  eine bessere Näherung herleitet. Dabei wird das lineare Optimierungsproblem

$$\hat{x}_k \in \operatorname{argmin}_{x \in D} \langle x, \nabla f(x_{k-1}) \rangle \quad (6.1)$$

gelöst, und mit  $\hat{x}_k$  der nächste Iterationsschritt konstruiert. An dieser Stelle existieren verschiedene Wege  $x_k$  zu berechnen. Doch bevor wir uns den Möglichkeiten zuwenden, eine

kurze Erklärung zu obigem Optimierungsproblem 6.1 und weshalb es linear ist. Durch Addition von konstanten Termen (bzw. von  $x \in D$  unabhängigen) erhalten wir den Zusammenhang

$$\begin{aligned} \operatorname{argmin}_{x \in D} \langle x, \nabla f(x_{k-1}) \rangle &= \operatorname{argmin}_{x \in D} f(x_{k-1}) + \langle x, \nabla f(x_{k-1}) \rangle - \langle x_{k-1}, \nabla f(x_{k-1}) \rangle \\ &= \operatorname{argmin}_{x \in D} f(x_{k-1}) + \langle x - x_{k-1}, \nabla f(x_{k-1}) \rangle =: \operatorname{argmin}_{x \in D} \widehat{f}(x) . \end{aligned}$$

Die Funktion  $\widehat{f}$  ist mit dem Skalarprodukt offensichtlich linear und daher eine Linearisierung von  $f$  an der Stelle  $x_{k-1}$ . Anders interpretiert entspricht  $\widehat{f}$  dem ersten Taylorpolynom von  $f$  an der Stelle  $x_{k-1}$ . Des Weiteren ist der Term  $\langle x, \nabla f(x_{k-1}) \rangle$ , über den minimiert wird, die Richtungsableitung von  $f$  am Punkt  $x_{k-1}$  in Richtung  $x \in D$ .

Für die Ermittlung von  $x_k$  bieten sich drei Methoden an. Als erstes, sei die von Frank-Wolfe selbst vorgestellte Konvexkombination erwähnt. Für  $\gamma := \frac{2}{k+2}$  wird der  $k$ -te Wert berechnet, indem die Lösung des linearen Minimierungsproblems und des zuvor erhaltenen  $x_{k-1}$  wie folgt in  $x_k$  eingehen:

$$x_k := (1 - \gamma)x_{k-1} + \gamma \widehat{x}_k .$$

Eine andere Möglichkeit bietet die sogenannte *Linienuche*, bei der  $x_k$  durch die Lösung eines neuen Minimierungsproblems ermittelt wird

$$x_k \in \operatorname{argmin}_{x \in \operatorname{conv}\{x_{k-1}, \widehat{x}_k\}} f(x) .$$

Zuletzt sei die *fully-corrective* Variante beschrieben. Im Gegensatz zu der Linienuche wird hier über die konvexe Hülle über alle zuvor ermittelten Lösungen  $\widehat{x}_{k-1}, \widehat{x}_{k-2}, \dots, \widehat{x}_1$  des linearen Minimierungsproblems minimiert:

$$x_k \in \operatorname{argmin}_{x \in \operatorname{conv}\{\widehat{x}_1, \dots, \widehat{x}_k\}} f(x) .$$

Verglichen zum klassischen Vorgehen durch das Bilden der Konvexkombination verspricht die fully-corrective Variante größeren Fortschritt pro Iteration und führt damit schneller zu einer Lösung  $x$ , die aus wenig Nicht-Null-Einträgen besteht (besser dünnbesetzt ist). Der Nachteil besteht jedoch darin, pro Iteration ein möglicherweise schweres Minimierungsproblem lösen zu müssen. Daher kann für die Methode keine Konvergenzrate angegeben werden. Im Gegensatz dazu ergibt sich für die Lösungssuche mittels der von Frank-Wolfe vorgeschlagenen Konvexkombination eine berechenbare Konvergenzrate. Hierfür sei  $f$   $\beta$ -glatt über  $D$  bezüglich einer Norm  $\|\cdot\|$  auf  $D$ , d.h. für alle  $x, y \in D$  gilt

$$f(y) \leq f(x) + \nabla f(x) \cdot (y - x) + \frac{\beta}{2} \|x - y\|^2 .$$

Weiter halten wir den Durchmesser von  $D$  bezüglich  $\|\cdot\|$  durch

$$\delta := \max_{x, y \in D} \|x - y\|$$

fest. Insgesamt gilt dann für eine Lösung  $x^* \in \underset{x \in D}{\operatorname{argmin}} f(x)$  und  $k \geq 1$

$$f(x_k) - f(x^*) \leq \frac{8\delta^2\beta}{k} = \mathcal{O}\left(\frac{1}{k}\right).$$

Für die zugehörigen Beweise und nähere Erläuterungen sei auf das Paper von Dan Garber und Elad Hazan [5] verwiesen.

## 6.2 Diskretisierung und die Richtungsableitung

Um dem unendlich-dimensionalen Minimierungsproblem numerisch handhabbar zu werden, muss eine Diskretisierung vorgenommen werden. Beginnend mit dem Tomographen definieren wir für  $s \in \mathbb{N}$  und der Anzahl an Messungen  $K \in \mathbb{N}$  das Gitter

$$T_{\#} := \left\{ \frac{1}{s \cdot K} j \in [0, 1]^d \mid j \in \{0, \dots, s \cdot K - 1, s \cdot K\}^d \right\} \subset [0, 1]^d.$$

Es gilt offensichtlich  $|T_{\#}| = (s \cdot K + 1)^d$ . Für den Phasenraum betrachten wir die Diskretisierung der Geschwindigkeitskomponenten. Im Gegensatz zu bisher setzen wir die Geschwindigkeiten nun abhängig von dem Zeitunterschied zwischen zwei Aufnahmen, also von  $t_k - t_{k-1} = \frac{T}{K}$ . Der Vorteil dieses Vorgehens wird später verdeutlicht. Innerhalb der Zeit dürfen sich die radioaktiven Zellen nun maximal mit betragsweise  $\frac{1}{K}$  in jede Richtung bewegen, ansonsten verlassen sie bereits vor dem Beenden der Messungen den Bereich des Tomographen. Dementsprechend sind die Geschwindigkeiten Elemente der Menge

$$V_{\#} := \left\{ \frac{1}{s \cdot K} \tilde{j} \in \left[ -\frac{1}{K}, \frac{1}{K} \right]^d \mid \tilde{j} \in \{-s, \dots, s\}^d \right\} \subset \left[ -\frac{1}{K}, \frac{1}{K} \right]^d.$$

Genauer erklärt haben wir die Messzeitpunkte  $t_k = k \frac{T}{K}$  und die Indizierung der Geschwindigkeiten mit  $t \in [0, T]$  durch  $k$  ausgetauscht, in dem wir die Konstante  $\frac{T}{K}$  bereits in die Geschwindigkeiten eingerechnet haben. Um die original vorliegende Geschwindigkeit zu erhalten, muss daher die rekonstruierte Geschwindigkeit mit  $\frac{K}{T}$  multipliziert werden. Insgesamt besitzt die diskrete Version des Phasenraumes folgende Form:

$$\Omega_{\#} := \{(x, v) \in T_{\#} \times V_{\#} \mid x + kv \in [0, 1]^d \text{ für alle } k \in \{0, \dots, K - 1\}\}.$$

Durch die Definition von  $\Omega_{\#}$  in Abhängigkeit von  $k$ , anstatt von  $t \in [0, T]$ , fällt somit der Länge des Messzeitraumes  $T$  bei der numerischen Lösung keine Bedeutung mehr zu. Ferner ergibt sich hier auch der Sinn hinter der Umstellung der Geschwindigkeiten. Dieser lässt sich erkennen, wenn wir die Position des Tupels  $(x, v) \in \Omega_{\#}$  zum Zeitpunkt der  $k$ -ten Messung bestimmen. Dafür sei  $x = \frac{1}{s \cdot K} j \in T_{\#}$  und  $v = \frac{1}{s \cdot K} \tilde{j} \in V_{\#}$ , mit zugehörigen Vektoren  $j, \tilde{j}$ . Dann gilt

$$x + kv = \frac{1}{s \cdot K} j + k \cdot \frac{1}{s \cdot K} \tilde{j} = \frac{j + k\tilde{j}}{s \cdot K} \in T_{\#}.$$

So ist garantiert, dass die Positionen aller Elemente aus  $\Omega_{\#}$  multipliziert mit ihren zugehörigen Geschwindigkeitskomponenten stets Punkte des diskretisierten Tomographen sind. Den diskreten Rand vom Tomographen definieren wir durch

$$\partial T_{\#} := \left\{ \frac{1}{s \cdot K} j \in T_{\#} \mid j \in \{0, \dots, s \cdot K - 1, s \cdot K\}^d, \right. \\ \left. j_i \in \{0, s \cdot K\} \text{ für mindestens eine Komponente } i \text{ von } j \right\} \subset \partial[0, 1]^d$$

und für den Abdeckungsbereich eines Detektorenpaares schreiben wir nach wie vor  $\Gamma_{(i)} \subset \partial T_{\#}$ , wobei  $\Gamma_{(i)}$  nun endlich viele diskrete Punkte beinhaltet. Aus Gründen der Leserlichkeit wird hier auf eine gesonderte Notation verzichtet. Für die Anzahl an Punkten des Randes von  $T_{\#}$  erhalten wir  $|\partial T_{\#}| = (s \cdot K + 1)^d - (s \cdot K - 1)^d$  (Anzahl der Punkte des Tomographen minus der Anzahl der inneren Punkte).

Durch die Endlichkeit des diskreten Phasenraumes ist nun jedes Radon-Maß auf  $\Omega_{\#}$  eine Linearkombination von Dirac-Maßen. Das heißt  $\lambda \in M(\Omega_{\#})$  hat die Form

$$\lambda = \sum_{a \in \Omega_{\#}} w_a \delta_a,$$

wobei  $w := (w_a)_{a \in \Omega_{\#}} \in \mathbb{R}^n$  der zum Maß gehörende Gewichtsvektor ist, mit  $n := |\Omega_{\#}|$ . Dieses Wissen möchten wir im nächsten Schritt anwenden, indem wir für das Minimierungsfunktional nur Maße  $\lambda \in M(\Omega_{\#})$  als Argument betrachten. Dabei gehen wir wie bisher bewährt vor und betrachten die Komponenten von  $J^{y,p}$  einzeln, beginnend mit der TV-Norm. Für diese ist bekannt, dass sie für Linearkombinationen von Dirac-Maßen gleich der Summennorm  $\|\cdot\|_1$  des zugehörigen Gewichtsvektors ist. Da zudem die vorliegenden Maße aus  $M(\Omega_{\#})$  per Definition nicht negativ sind, wird folgender Zusammenhang ersichtlich:

$$\|\lambda\|_{TV} = \|w\|_1 = \sum_{a \in \Omega_{\#}} |w_a| = \sum_{a \in \Omega_{\#}} w_a.$$

Aufgrund der vorgenommenen Skalierung der Geschwindigkeiten mittels  $\frac{T}{K}$  ändern sich die Messintervalle  $[t_k, t_{k+1}]$  hin zu  $[k, k+1]$ , für  $k \in \{0, \dots, K-1\}$ . Seien  $k$  und  $i$  beliebig, dann ergibt sich

$$\begin{aligned} C_{k,i}^S(\lambda) &= p_S \int_k^{k+1} (\phi_t)_{\#} \lambda(T_{\#}) \cdot \frac{|\Gamma_{(i)}|}{|\partial T_{\#}|^2} dt \\ &= p_S \int_k^{k+1} \lambda(\Omega_{\#}) \frac{|\Gamma_{(i)}|}{|\partial T_{\#}|^2} dt \\ &= \frac{p_S |\Gamma_{(i)}|}{((s \cdot K + 1)^d - (s \cdot K - 1)^d)^2} \int_k^{k+1} \sum_{a \in \Omega_{\#}} w_a \delta_{(a)}(\Omega_{\#}) dt \\ &= \frac{p_S |\Gamma_{(i)}|}{((s \cdot K + 1)^d - (s \cdot K - 1)^d)^2} \|w\|_1. \end{aligned}$$

Wir notieren im Folgenden für den von der Dimension, dem Detektorenpaar und der Anzahl an Messungen abhängigen Term

$$c_{K,d,i} := \frac{p_S |\Gamma_{(i)}|}{((s \cdot K + 1)^d - (s \cdot K - 1)^d)^2}.$$

Selbstverständlich spielen weitere Abhängigkeiten in  $c_{K,d,i}$  eine Rolle, wie die Wahl der Feinheit des Gitters durch  $s$  und die Streuungskonstante  $p_S$ . Jedoch begnügen wir uns hier mit der Indizierung über  $K$ ,  $d$  und  $i$ . Der Term  $c_{K,d,i}$  verdeutlicht nochmals den Gedanken hinter dem die Streuung modellierenden Teil des Rekonstruktionsfunktional. Er gibt, unter Nicht-Beachtung von  $p_S$ , den Anteil vom  $i$ -ten Detektorenpaar am ganzen Abdeckungsbereiche wieder. Für  $C_{k,i}^E$  erhalten wir

$$\begin{aligned} C_{k,i}^E(\lambda) &= p_E \int_k^{k+1} \int_{G(1,d)} \int_{P_z^{-1}(\Gamma_{(i)},v)} \int_{[0,1]^d} G_x(y) d(\phi_t)_\# \lambda(x) dy dv_d dt \\ &= p_E \int_k^{k+1} \int_{G(1,d)} \int_{P_z^{-1}(\Gamma_{(i)},v)} \sum_{a \in \Omega_\#} w_a \cdot G_{\phi_t(a)}(y) dy dv_d dt \\ &= p_E \int_k^{k+1} \int_{G(1,d)} \sum_{y \in P_z^{-1}(\Gamma_{(i)},v)} \sum_{a \in \Omega_\#} w_a \cdot G_{\phi_t(a)}(y) dv_d dt. \end{aligned}$$

Interpretieren wir  $G(1,d)$  im diskreten als Menge aller Richtungen von Linien, die zwischen jeweils zwei Randpunkten von  $T_\#$  gezogen werden können, so folgt:

$$\begin{aligned} C_{k,i}^E(\lambda) &= p_E \int_k^{k+1} \sum_{v \in G(1,d)} \sum_{y \in P_z^{-1}(\Gamma_{(i)},v)} \sum_{a \in \Omega_\#} w_a \cdot G_{\phi_t(a)}(y) dt \\ &= p_E \sum_{v \in G(1,d)} \sum_{y \in P_z^{-1}(\Gamma_{(i)},v)} \sum_{a \in \Omega_\#} \int_k^{k+1} w_a \cdot G_{\phi_t(a)}(y) dt. \end{aligned}$$

Der Gleichung liegt zu Grunde, dass das Haar-Maß für diskrete Mengen dem Zählmaß entspricht. Für die Auswertung des Integrals nutzen wir die Mittelpunktsregel, so dass wir für einen einzelnen Summanden folgendes schreiben können

$$\begin{aligned} \int_k^{k+1} w_a \cdot G_{\phi_t(a)}(y) dt &\approx w_a \cdot ((k+1) - k) \cdot G_{\phi_{(2k+1)/2}(a)}(y) \\ &= w_a \cdot G_{(\phi_{k+1}(a) + \phi_k(a))/2}(y) \end{aligned}$$

Wir nutzen den Umweg über die Rechnung  $(\phi_{k+1}(a) + \phi_k(a))/2$ , da  $\phi_k(a)$  nur für ganzzahlige  $k$  Element des diskretisierten Tomographen  $T_\#$  ist. Fassen wir nun alles zusammen, so gilt

insgesamt

$$\begin{aligned}
 J^{y,p}(\lambda) &= \sum_{k=0}^{K-1} \sum_{i=1}^{\widehat{D}} \left[ C_{k,i}(\lambda) - y_{k,i} \log(p \cdot C_{k,i}^S(\lambda) + C_{k,i}^E(\lambda)) \right] + \beta \|\lambda\|_{TV} \\
 &\approx \sum_{k=0}^{K-1} \sum_{i=1}^{\widehat{D}} \left[ c_{K,d,i} \|w\|_1 + p_E \sum_{v \in G(1,d)} \sum_{y \in P_z^{-1}(\Gamma_{(i),v})} \sum_{a \in \Omega_{\#}} w_a \cdot G_{(\phi_{k+1}(a) + \phi_k(a))/2}(y) \right. \\
 &\quad \left. - y_{k,i} \log \left( p \cdot c_{K,d,i} \|w\|_1 + p_E \sum_{v \in G(1,d)} \sum_{y \in P_z^{-1}(\Gamma_{(i),v})} \sum_{a \in \Omega_{\#}} w_a \cdot G_{(\phi_{k+1}(a) + \phi_k(a))/2}(y) \right) \right] \\
 &\quad + \beta \|w\|_1 .
 \end{aligned}$$

Es liegt nahe, eine Funktion  $J_{dis}^{y,p}$  zu definieren, die als Argument  $|\Omega_{\#}|$ -dimensionale Vektoren  $w$  benutzt (um Verwechslungen mit dem Bildmaß zu vermeiden, schreiben wir hier *dis*, anstatt die Raute zu verwenden)

$$\begin{aligned}
 J_{dis}^{y,p} : \mathbb{R}_+^n &\longrightarrow \mathbb{R} \\
 w &\longmapsto J_{dis}^{y,p}(w) := J^{y,p}(\lambda) \quad \text{mit } \lambda = \sum_{a \in \Omega_{\#}} w_a \delta_a .
 \end{aligned} \tag{6.2}$$

Für die numerische Lösung des Minimierungsproblems 4.10, können wir folglich auf das endlich-dimensionale Problem

$$\min_{w \in \mathbb{R}_+^n} J_{dis}^{y,p}(w) \tag{6.3}$$

zurückgreifen. Durch diese Umformulierung nähern wir uns einer Struktur, die es uns ermöglicht, den Frank-Wolfe-Algorithmus anzuwenden. Als dessen Bestandteil gilt es, die Richtungsableitung von  $J_{dis}^{y,p}$  herzuleiten. Da  $J_{dis}^{y,p}(w)$  für alle  $w$  total differenzierbar ist, schreiben wir die Richtungsableitung mittels des Gradienten. An der Stelle 0 sei darauf hingewiesen, dass nur positive Maße betrachtet werden und die Ableitung bei 0 von rechts kommend stetig fortgeführt wird, wodurch die Differenzierbarkeit der 1-Norm bei 0 gegeben ist. Des Weiteren sei angemerkt, dass mit der gleichen Begründung die TV-Norm für positive Maße, und damit auch insgesamt  $J^{y,p}$ , Gâteaux-differenzierbar ist. Hier belassen wir es jedoch bei der Herleitung des für die Numerik benötigten Gradienten der diskreten Form  $J_{dis}^{y,p}$ . Für beliebige positive Vektoren  $w, v \in \mathbb{R}_+^n$  ungleich dem Nullvektor gilt

$$\nabla_v J_{dis}^{y,p}(w) = \langle \nabla J_{dis}^{y,p}(w), v \rangle = \sum_{a \in \Omega_{\#}} \frac{\partial}{\partial w_a} J_{dis}^{y,p}(w) v_a .$$

Betrachten wir einen einzelnen Summanden  $b \in \Omega_{\#}$  und beziehen die Rechnung  $\frac{\partial}{\partial w_b} \|w\|_1 = 1$  und die Kettenregel ein, so ergibt sich

$$\begin{aligned}
 & \frac{\partial}{\partial w_b} J_{dis}^{y,p}(w) v_b \\
 & \approx v_b \sum_{k=0}^{K-1} \sum_{i=1}^{\widehat{D}} \left[ c_{K,d,i} \frac{\partial}{\partial w_b} \|w\|_1 + p_E \sum_{v \in G(1,d)} \sum_{y \in P_z^{-1}(\Gamma_{(i),v})} \sum_{a \in \Omega_{\#}} \frac{\partial}{\partial w_b} w_a \cdot G_{(\phi_{k+1}(a)+\phi_k(a))/2}(y) \right. \\
 & \quad \left. - y_{k,i} \frac{p \cdot c_{K,d,i} \frac{\partial}{\partial w_b} \|w\|_1 + p_E \sum_{v \in G(1,d)} \sum_{y \in P_z^{-1}(\Gamma_{(i),v})} \sum_{a \in \Omega_{\#}} \frac{\partial}{\partial w_b} w_a G_{(\phi_{k+1}(a)+\phi_k(a))/2}(y)}{p \cdot c_{K,d,i} \|w\|_1 + p_E \sum_{v \in G(1,d)} \sum_{y \in P_z^{-1}(\Gamma_{(i),v})} \sum_{a \in \Omega_{\#}} w_a G_{(\phi_{k+1}(a)+\phi_k(a))/2}(y)} \right] \\
 & + v_b \cdot \beta \frac{\partial}{\partial w_b} \|w\|_1 \\
 & = v_b \sum_{k=0}^{K-1} \sum_{i=1}^{\widehat{D}} \left[ c_{K,d,i} + p_E \sum_{v \in G(1,d)} \sum_{y \in P_z^{-1}(\Gamma_{(i),v})} G_{(\phi_{k+1}(b)+\phi_k(b))/2}(y) \right. \\
 & \quad \left. - y_{k,i} \frac{p \cdot c_{K,d,i} + p_E \sum_{v \in G(1,d)} \sum_{y \in P_z^{-1}(\Gamma_{(i),v})} G_{(\phi_{k+1}(b)+\phi_k(b))/2}(y)}{p \cdot c_{K,d,i} \|w\|_1 + p_E \sum_{v \in G(1,d)} \sum_{y \in P_z^{-1}(\Gamma_{(i),v})} \sum_{a \in \Omega_{\#}} w_a \cdot G_{(\phi_{k+1}(a)+\phi_k(a))/2}(y)} \right] \\
 & + v_b \beta .
 \end{aligned}$$

Durch das Summieren über alle  $b \in \Omega_{\#}$  erhalten wir schlussendlich die gesuchte Richtungsableitung

$$\begin{aligned}
 \nabla_v J_{dis}^{y,p}(w) & = \langle \nabla J_{dis}^{y,p}(w), v \rangle = \sum_{b \in \Omega_{\#}} \frac{\partial}{\partial w_b} J_{dis}^{y,p}(w) v_b \\
 & \approx \sum_{b \in \Omega_{\#}} \sum_{k=0}^{K-1} \sum_{i=1}^{\widehat{D}} \left[ c_{K,d,i} \cdot v_b + p_E \sum_{v \in G(1,d)} \sum_{y \in P_z^{-1}(\Gamma_{(i),v})} v_b \cdot G_{(\phi_{k+1}(b)+\phi_k(b))/2}(y) \right. \\
 & \quad \left. - y_{k,i} \frac{p \cdot c_{K,d,i} \cdot v_b + p_E \sum_{v \in G(1,d)} \sum_{y \in P_z^{-1}(\Gamma_{(i),v})} v_b \cdot G_{(\phi_{k+1}(b)+\phi_k(b))/2}(y)}{p \cdot c_{K,d,i} \|w\|_1 + p_E \sum_{v \in G(1,d)} \sum_{y \in P_z^{-1}(\Gamma_{(i),v})} \sum_{a \in \Omega_{\#}} w_a \cdot G_{(\phi_{k+1}(a)+\phi_k(a))/2}(y)} \right] \\
 & + \beta \sum_{b \in \Omega_{\#}} v_b \\
 & = \sum_{k=0}^{K-1} \sum_{i=1}^{\widehat{D}} \left[ c_{K,d,i} \cdot \|v\|_1 + p_E \sum_{v \in G(1,d)} \sum_{y \in P_z^{-1}(\Gamma_{(i),v})} \sum_{b \in \Omega_{\#}} v_b \cdot G_{(\phi_{k+1}(b)+\phi_k(b))/2}(y) \right. \\
 & \quad \left. - y_{k,i} \frac{p \cdot c_{K,d,i} \|v\|_1 + p_E \sum_{v \in G(1,d)} \sum_{y \in P_z^{-1}(\Gamma_{(i),v})} \sum_{b \in \Omega_{\#}} v_b \cdot G_{(\phi_{k+1}(b)+\phi_k(b))/2}(y)}{p \cdot c_{K,d,i} \|w\|_1 + p_E \sum_{v \in G(1,d)} \sum_{y \in P_z^{-1}(\Gamma_{(i),v})} \sum_{a \in \Omega_{\#}} w_a \cdot G_{(\phi_{k+1}(a)+\phi_k(a))/2}(y)} \right] \\
 & + \beta \|v\|_1 .
 \end{aligned}$$

Um den Frank-Wolfe-Algorithmus für  $J_{dis}^{y,p}$  nutzen zu können, werfen wir einen Blick auf die Menge, über die das Minimum gebildet wird. Bis hierhin entspricht unsere Herleitung

nicht den Voraussetzungen, denn  $\mathbb{R}_+^n$  ist keine kompakte Menge. Abhilfe verschafft folgende Definition:

$$W := \left\{ w \in \mathbb{R}_+^n \mid \|w\|_1 \leq \|\widehat{\lambda}\|_{TV} \right\}.$$

Hierbei ist  $\widehat{\lambda} \in M(\Omega)$  das original vorliegende, zu rekonstruierende Maß der radioaktiven Isotope. Vergegenwärtigen wir uns an dieser Stelle, dass in der Anwendung eine beschränkte (möglichst geringe) Anzahl an Isotopen verabreicht wird, so folgt sofort die Beschränktheit von  $W$  und insgesamt die Kompaktheit, ebenso wie die Konvexität. Das zu lösende Minimierungsproblem ist somit

$$\min_{w \in W} J_{dis}^{y,p}(w). \quad (6.4)$$

Abschließend sei auf eine nicht zu ändernde Tatsache hingewiesen: Leider scheitert im Allgemeinen eine exakte Rekonstruktion daran, dass  $\widehat{\lambda} \notin M(\Omega_{\#})$  gilt. Anders ausgedrückt ist die Menge aller wahren Positionen und Geschwindigkeiten  $\{(x_i, v_i) \in \Omega \mid i = 1, \dots, N\}$  keine Teilmenge des diskreten Phasenraumes. Die durch das Lösen von Gleichung 6.4 erhaltene Konfiguration der Positionen und Geschwindigkeiten entspricht also der tatsächlich Vorliegenden nur bis zu einem Fehler. Dieser Fehler ist abhängig von der Feinheit der Diskretisierung, bei uns von  $s \in \mathbb{N}$ . Je größer  $s$ , desto geringer fällt der Fehler aus. In der Mathematik ist dies auch unter dem Begriff *basis mismatch* bekannt.

### 6.3 Die Implementierung und Simulation der Messungen

Für die Lösung des Minimierungsproblems 6.4 greifen wir auf die zuerst vorgestellte Variante des Frank-Wolfe-Algorithmus zurück: Für die Ermittlung einer Näherung wird pro Iterationsschritt eine Konvexkombination zwischen vorheriger Näherung und der Lösung des zugehörigen linearen Problems 6.1 berechnet.

---

#### Algorithm 1 Frank-Wolfe-Algorithmus

---

```

for  $k = 1, \dots, k_{max}$  do                                ▷ Maximale Anzahl an Iterationsschritten:  $k_{max} \in \mathbb{N}$ 
     $\gamma_k = \frac{2}{2+k}$                                        ▷ Konvexkombinationskonstante
     $\widehat{w}_k = \underset{w \in W}{\operatorname{argmin}} \langle \nabla J_{dis}^{y,p}(w_{k-1}), v \rangle$    ▷ Lösen des linearen Minimierungsproblems
     $w_k = w_{k-1} + \gamma_k \cdot (\widehat{w}_k - w_{k-1})$            ▷ Neue Näherung
    if  $|J_{dis}^{y,p}(w_k) - J_{dis}^{y,p}(w_{k-1})| < \text{Toleranz}$  then   ▷ Berechnung der Verbesserung
         $w_k$  ist die Lösung
        Vorzeitiger Abbruch des Verfahrens
    end if
end for

```

---

Während die Implementierungen der meisten Bestandteile des Minimierungsfunktionals 6.2 selbsterklärend sind, bedarf es für die Menge  $P_z^{-1}(\Gamma_{(i)}, v)$  einer näheren Erläuterung. Hierfür

rufen wir uns die Funktion  $P$  in Erinnerung

$$P : [0, 1]^d \times G(1, d) \longrightarrow \partial[0, 1]^d \\ (z, v) \longmapsto (s_1, s_2) .$$

Durch Betrachtung der Umkehrfunktion mit ausschließlich Punkten aus  $\partial T_{\#}$  können wir die ursprüngliche Definition beibehalten und benötigen keine gesonderte diskrete Formulierung. Stattdessen gilt es, für die Menge

$$P_z^{-1}(\Gamma_{(i)}, v) = \{z \in [0, 1]^d \mid (z, v) \in P^{-1}(\Gamma_{(i)})\}$$

eine diskrete Formulierung zu finden, sodass

$$P_z^{-1}(\Gamma_{(i)}, v) \underset{\text{diskretisiert}}{=} \{z \in T_{\#} \mid (z, v) \in P^{-1}(\Gamma_{(i)})\}$$

nur Gitterpunkte des diskreten Tomographen beinhaltet. Aus Gründen der Leserlichkeit wird hier, ebenso wie bei  $\Gamma_{(i)}$ , auf eine gesonderte Notation verzichtet.

In  $P_z^{-1}(\Gamma_{(i)}, v)$  sind nach Definition von  $P$  alle Punkte des Tomographen enthalten, die auf einer durch  $v \in G(1, d)$  charakterisierten Linie zwischen den Detektoren des Paares  $i$  liegen. Folglich wird für die Implementierung im ersten Schritt die besagte Linie gebildet, um im zweiten Schritt alle Gitterpunkte aus  $T_{\#}$  zu finden, deren Abstand zur Linie unter einer gewissen Toleranzgrenze liegen. Eben diese Punkte werden der Menge  $P_z^{-1}(\Gamma_{(i)}, v)$  zugeschrieben.

Die Toleranzgrenze sollte in enger Verbindung mit der Feinheit des Gitters stehen, genauer bietet es sich an sie abhängig vom Mindestabstand zwischen zwei benachbarten Gitterpunkten aus  $T_{\#}$  zu wählen. Dass überhaupt mit einer Toleranzgrenze gearbeitet werden muss, liegt an der Diskretisierung des Tomographen, schließlich verlaufen die konstruierten Linien  $v$  in  $[0, 1]^d$  und können daher an sämtlichen inneren Punkte von  $T_{\#}$  vorbeiführen.

Diesem Setting nach entspricht  $P_z^{-1}(\Gamma_{(i)}, v)$  in einem gewissen Maße der bei Abschnitt 3.4 eingeführten *line of response*. Jedoch wird hier nur die *line of response* einer einzigen Richtung betrachtet, die *line of response* eines Detektorpaares  $i$  erhalten wir mit

$$\bigcup_{v \in G(1, d)} P_z^{-1}(\Gamma_{(i)}, v) .$$

Für eine detailreiche Beschreibung der Implementierung des Gradienten von  $J_{dis}^y$  sei auf den Anhang verwiesen. Die Simulation von  $N \in \mathbb{N}$  Isotopen in  $T_{\#}$ , mit zugehörigen Positionen und Geschwindigkeiten, wird durch die Bestimmung  $N$  zufälliger Elemente aus  $\Omega_{\#}$  vorgenommen. Den zu der so erhaltenen Teilchenkonfiguration  $\{(x_i, v_i) \in \Omega_{\#} \mid i = 0, \dots, N\}$  passenden Vektor  $\hat{w}$  konstruieren wir durch

$$\hat{w} = (\hat{w}_a)_{a \in \Omega_{\#}} ,$$

wobei für die einzelnen Komponenten  $a \in \Omega_{\#}$  gilt

$$\hat{w}_a = \begin{cases} 1 & \text{falls } a \in \{(x_i, v_i) \in \Omega_{\#} \mid i = 0, \dots, N\} \\ 0 & \text{sonst} \end{cases} .$$

Auch wenn diese Definition auf den ersten Blick richtig erscheint, so gilt es noch eine entscheidende Umformulierung vorzunehmen. In Abschnitt 3.4 haben wir die Zerfallskonstante invers mit dem Zielmaß multipliziert, um für den Verlauf der Arbeit auf die Notationen innerhalb des Rekonstruktionsfunktionalen verzichten zu können. Demnach gilt es auch  $\hat{w}$  mit  $\frac{T_{1/2}}{\ln 2}$  zu multiplizieren und der Vektor, der die Teilchenkonfiguration repräsentiert, ist folglich

$$\frac{T_{1/2}}{\ln 2} \cdot \hat{w}.$$

Ebenso haben wir den Regularisierungsparameter  $\beta$  mit der Zerfallskonstanten skaliert, so dass der ursprüngliche Regularisierungsparameter  $\tilde{\beta} = \beta \cdot \frac{\ln 2}{T_{1/2}}$  entspricht. Der nächste Schritt besteht darin, die gemessenen Daten  $y$  zu simulieren, sprich die Anzahl an Detektionen eines jeden Detektorpaares  $i$  für alle Zeitpunkte  $k$ . Wie erneut in Abschnitt 3.4 beschrieben, entsprechen die Messungen der Realisierung einer  $\text{Poi}(C_{k,i}(\hat{\lambda}))$  verteilten Zufallsvariable. Das Maß  $\hat{\lambda}$  erhalten wir mittels der Identifizierung über  $\frac{T_{1/2}}{\ln 2} \cdot \hat{w}$

$$\hat{\lambda} = \sum_{a \in \Omega_{\#}} \frac{T_{1/2}}{\ln 2} \cdot \hat{w}_a \delta_a$$

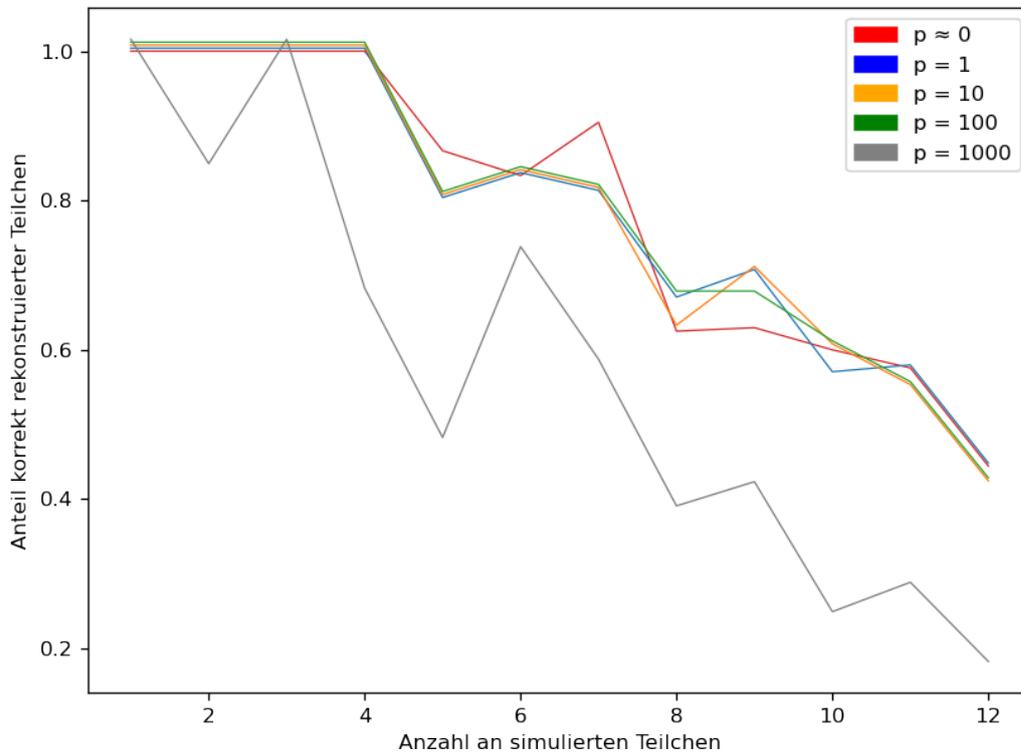
und somit sind die Messungen durch

$$y = (y_{k,i})_{k,i} = (Y_{k,i}(\tilde{\omega}))$$

gegeben, mit  $Y_{k,i}$  als  $\text{Poi}(C_{k,i}(\hat{\lambda}))$  verteilter Zufallsvariable ( $\tilde{\omega} \in \tilde{\Omega}$ , wobei  $\tilde{\Omega}$  den zum Poisson-Punkt-Prozess gehörenden Ereignisraum definiert).

## 6.4 Numerische Ergebnisse

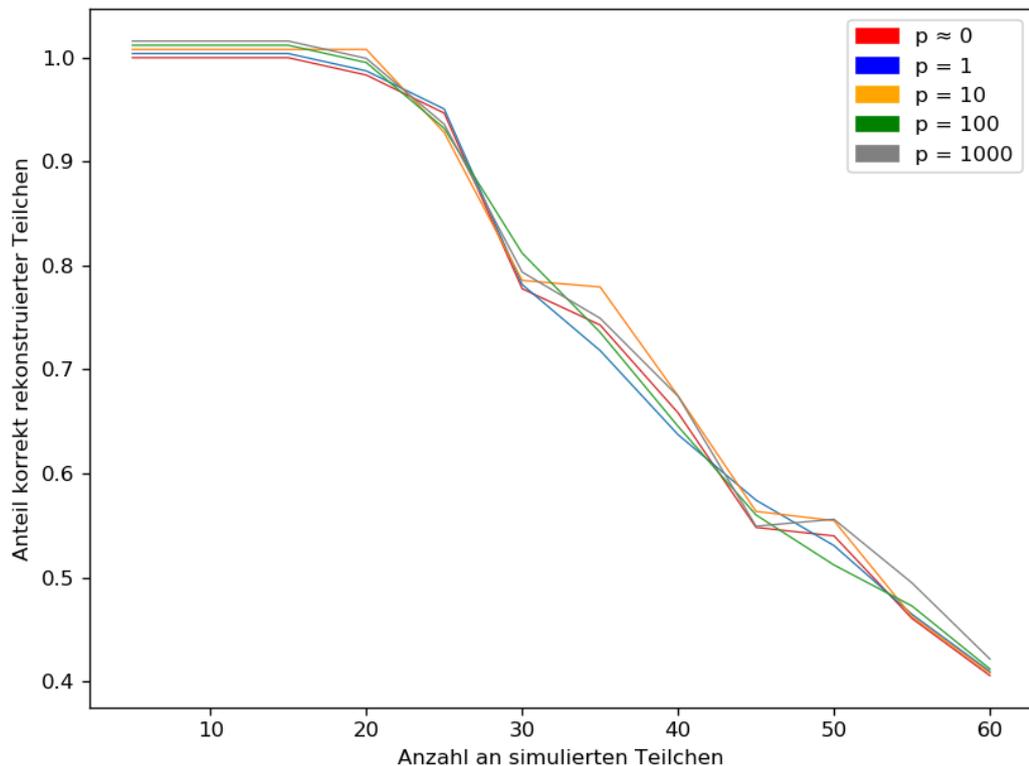
Wir untersuchen im Folgenden die Genauigkeit des Rekonstruktionsverfahrens im Zweidimensionalen für verschiedene Werte des Streuungsparameters  $p$  und verschiedene Mengen an zu simulierenden Teilchen. Hierfür setzen wir die Wahrscheinlichkeit der Streuung,  $p_S$ , auf 0.2, die Wahrscheinlichkeit  $p_E$  (das Photonennpaar wird ohne Streuung detektiert) setzen wir auf 0.75. Somit führt jeder zwanzigste Zerfall nicht zu einer Messung. Wir setzen den Regularisierungsparameter  $\beta$  auf eins und die Standardabweichung  $\epsilon$  abhängig von dem Abstand benachbarter Gitterpunkte auf  $\frac{1}{s \cdot K}$ . Weiter simulieren wir die Rekonstruktion zunächst auf einem  $10 \times 10$  Gitter, definiert durch  $K = 3$  Messungen,  $s = 3$  und zwölf Detektoren, anschließend betrachten wir ein  $31 \times 31$  Gitter mit  $K = 5$ ,  $s = 6$  und 30 Detektoren. Als radioaktive Teilchen betrachten wir für das  $10 \times 10$  Gitter bis zu zwölf Isotope  $^{18}\text{F}$  von Fluor, die jeweils eine Halbwertszeit von  $T_{1/2} = 110$  Minuten besitzen, für das  $31 \times 31$  simulieren wir in in zwölf Schritten bis zu 60 Isotope. Jede Anzahl der Teilchen wird für drei zufällige Anordnungen simuliert, um anschließend den Anteil der durch das Rekonstruktionsverfahren korrekt ermittelten Teilchen anzugeben.

Abbildung 6.1:  $10 \times 10$  Gitter

In der obigen Grafik ist die Auswertung des Rekonstruktionsverfahrens auf dem  $10 \times 10$  Gitter mit fünf verschiedenen Streuungsparametern zu sehen. Augenscheinlich steht die Genauigkeit aller Ansätze in starker Abhängigkeit zu der Anzahl an zu simulierenden Teilchen. Während für geringe Mengen an Fluor alle Verfahren vielversprechend genau die Positionen und Geschwindigkeiten rekonstruieren, nehmen mit steigender Menge Fehlrekonstruktionen zu. Dies ist darauf zurückzuführen, dass mit steigender Anzahl auch der Abstand zwischen zwei Isotopen geringer wird, wodurch es dem Verfahren erschwert wird exakt zu rekonstruieren (siehe Candès et al. in [7] und Schlueter in [20]).

Interessant ist der Vergleich zwischen den verschiedenen Streuungsparametern. Während mit  $p \approx 0$  eine Rekonstruktion vorgenommen wird, bei der nahezu alle Messereignisse als nicht-gestreut interpretiert werden, verhält sich das Verfahren mit  $p = 1000$  konträr und legt die meisten Ereignisse als gestreut aus. Wie in der Grafik zu sehen, ist die Rekonstruktion mit einer derart starken Gewichtung des Streuungsterms für alle Anzahlen an simulierten Teilchen ungenauer, als durch moderat gewählte Parametergrößen. Diese Erkenntnis erscheint logisch, da nach Konstruktion von  $C_{k,i}^S$  die Intensität eines jeden Zerfalls gleichmäßig auf alle Detektorenpaare verteilt wird und der relative Messunterschied zwischen den

Detektorenpaaren nichtig wird. Offensichtlich verletzt damit  $p = 1000$  eine der beiden an  $p$  gesetzten Voraussetzungen 4.6 oder 4.7. Entgegen der Erwartung fällt eine Rekonstruktion durch  $p$  nahe 0 lediglich durch stärkere Rekonstruktionsschwankungen auf. Insgesamt scheint es aber eine geeignete Wahl zu sein, für manche Teilchenkonfigurationen sogar die Beste. Das Rekonstruktionsverfahren mit den Streuungsparametern 1, 10 und 100 ermittelt nahezu identische Ergebnisse, woraus zu schließen ist, dass alle drei Werte für die betrachteten Messungen den Voraussetzungen 4.6 und 4.7 genügen.

Abbildung 6.2:  $31 \times 31$  Gitter

Betrachten wir das  $31 \times 31$  Gitter, so fallen im Vergleich zu vorher zwei wesentliche Unterschiede auf: Zum einen werden hier bis zu 15, für  $p = 10$  sogar 20, Teilchen exakt erfasst. Zum anderen konstruieren die verschiedenen Streuungsparameter ähnliche Ergebnisse, der Wert  $p = 1000$  fällt nicht aus dem Raster. Während die exakte Rekonstruktion von mehr Teilchen auf einem größeren Gitter eine logische Schlussfolgerung ist, da die Abstände zwischen den Teilchen erst für größere Mengen zu klein werden, zeigt sich für den Streuungsparameter, dass dieser ebenfalls abhängig von der Wahl der Gittergröße ist. Dies wird mathematisch auch in den Ungleichungsvoraussetzungen 4.6 und 4.7 durch die Abhängigkeit des Parameters von  $K$  und  $\hat{D}$  deutlich.

## 7 Ausblick

Innerhalb dieser Arbeit wurde ein medizinisches Bildgebendes Verfahren vorgestellt, das die verschiedenen Pfade von radioaktiv markierten Zellen innerhalb eines Untersuchungsobjekts rekonstruiert.

Dabei lag das Augenmerk zunächst auf der Modellierung der statischen Rekonstruktion und der Vorstellung des Phasenraumes. Anschließend wurden die erworbenen Kenntnisse für die Herleitung eines dynamischen Messoperators genutzt, um im nächsten Schritt das nach Alberti et al. zur Rekonstruktion gehörende Minimierungsproblem aufzustellen. Zur Umgehung der Schlechtgestellttheit des Minimierungsproblems griffen wir auf eine Maximum-a-posteriori Schätzung zurück, das dadurch aufgestellte Rekonstruktionsfunktional erwies sich jedoch nicht als zielführend, da es Neigungen zeigte Detektionen als primär nicht-gestreute Ereignisse zu interpretieren. Um dem entgegenzutreten führten wir einen Streuungsparameter ein, des Weiteren formulierten wir darauf aufbauend eine konvex relaxierte Version des Rekonstruktionsfunktional.

Dieses analysierten wir anschließend hinsichtlich auf die Existenz von Lösungen. Es stellte sich heraus, dass Minimierer existieren, jedoch kein Rückschluss auf deren Eindeutigkeit zu ziehen ist, da das Funktional lediglich konvex und nicht strikt konvex ist. Als Leitfaden für den Beweis der Existenz nutzten wir die in den Grundlagen vorgestellte direkte Methode.

Den Abschluss der Arbeit bildete die numerische Simulation des Verfahrens, wobei wir die Rekonstruktion mit verschiedenen Werten des Streuungsparameters und verschiedene Anzahlen an vorliegenden Isotopen vollzogen haben. Als ein entscheidendes Ergebnis ließ sich ablesen, dass die Funktionsfähigkeit des Verfahrens stark im Zusammenhang zu der Anzahl an zu simulierenden Teilchen steht, unabhängig von der Wahl des Streuungsparameters. Dies liegt an dem mit wachsender Anzahl zu geringen Abstand zwischen den Teilchen, der einer exakten Rekonstruktion hinderlich ist.

Die vom Algorithmus rekonstruierten Maße beinhalteten als Summe über Dirac-Maße Daten über die Startposition und Geschwindigkeitsvektoren einer jeden Zelle, wodurch sich die zugehörigen linearen Bewegungen ablesen lassen. An dieser Stelle sticht die starke Einschränkung durch die Linearität hervor.

Die Annahme konnte nur durch die Betrachtung möglichst kurzer Messzeiträume gewährleistet werden. Diese haben jedoch den Nachteil, dass pro Messintervall wiederum wenige gemessene Zerfälle als Datensatz zur Verfügung stehen. Für genauere Ergebnisse muss auf höher radioaktivere Stoffe zurückgegriffen werden, was es jedoch in der Anwendung stets unter Beachtung der Gesundheit des Untersuchungsobjektes zu vermeiden gilt. Diese Austauschbeziehung genauer auf das Für und Wider hin zu analysieren, wie auch die PET-Bildrekonstruktion im Phasenraum für kurvige Pfade anwendbar zu modellieren, sind offene

Forschungsfragen. Hinsichtlich der hohen Dimensionalität des Phasenraum-Modells gilt es ebenso das Rekonstruktionsfunktional mittels dimensionsreduzierenden Verfahren handhabbarer zu gestalten. All diese Fragestellungen sind als Anreiz zu verstehen, sich weiter mit dem spannenden Thema der PET-Bildrekonstruktion im Phasenraum zu beschäftigen.

# Literatur

- [1] Antonin Chambolle et. al. *An introduction to Total Variation for Image Analysis*. 2009. URL: <https://hal.archives-ouvertes.fr/hal-00437581>.
- [2] Benedikt Wirth. *Inverse Probleme, Vorlesung SS 2019, Münster*.
- [3] Benedikt Wirth, Klaus P. Schäfers, Bernhard Schmitzer. *Dynamic Cell Imaging in PET with Optimal Transport Regularization*. 2019. URL: <https://arxiv.org/abs/1902.07521>.
- [4] Martin Burger. *Inverse Probleme, Vorlesung WS 16-17, Münster*. 2017.
- [5] Dan Garber, Elad Hazan. *Faster Rates for the Frank-Wolfe Method over Strongly-Convex Sets*. 2015. URL: <https://arxiv.org/abs/1406.1305>.
- [6] Jürgen Elstrodt. *Maß- und Integrationstheorie*. Springer, 2011. ISBN: 978-3-642-17905-1.
- [7] Emmanuel J. Candès, Carlos Fernandez-Granda. *Towards a Mathematical Theory of Super-Resolution*. 2012.
- [8] Gerald B. Folland. *Real Analysis: Modern Techniques and Their Applications*. John Wiley und Sons, 1999. ISBN: 978-0-471-31716-6.
- [9] Giovanni Alberti. *Dynamic Spike Super-resolution and Applications to Ultrafast Ultrasound Imaging*. 2018. URL: <https://arxiv.org/abs/1803.03251>.
- [10] Günther Last, Mathew Penrose. *Lectures on the Poisson Process*. 2017.
- [11] Irene Fonseca, Giovanni Leoni. *Modern Methods in the Calculus of Variations: Lp Spaces*. Springer Science+Business Media, 2007. ISBN: 978-0-387-69006-3.
- [12] Martin Jaggi. *Revisiting Frank-Wolfe: Projection-Free Sparse Convex Optimization*. 2013.
- [13] Jorge Nocedal, Stephen J. Wright. *Numerical Optimization*. Springer, 2006. ISBN: 978-0387-30303-1.
- [14] Keum Sil Lee, Tae Jin Kim, Guillem Pratx. *Single-Cell Tracking With PET Using a Novel Trajectory Reconstruction Algorithm*. Apr. 2015. DOI: 10.1109/TMI.2014.2373351.
- [15] Kristian Bredies, Hanna Katriina Pikkarainen. *INVERSE PROBLEMS IN SPACES OF MEASURES*. 2012. DOI: 10.1051/cocv/2011205.
- [16] Marco Jonas Mauritz. „Analysis of a Dynamic Cell Imaging Model in Positron Emission Tomography“. Magisterarb. University of Münster, 2020.

- [17] Marguerite Frank, Philip Wolfe. *An algorithm for quadratic programming*. Naval Research Logistics Quarterly, 1956. DOI: 10.1002/nav.3800030109.
- [18] Rolf Schneider, Wolfgang Weil. *Stochastic and Integral Geometry*. Springer, 2008. ISBN: 978-3-540-78858-4.
- [19] Walter Rudin. *REAL AND COMPLEX ANALYSIS*. Hrsg. von Peter R. Devine. 3. Aufl. McGraw-Hill Book Company, 1987. ISBN: 0-07-100276-6.
- [20] Alexander Schlueter. „Dimensionsreduktion von dynamischer Super-Resolution“. Masterarbeit. University of Münster, 2019.
- [21] Dirk Werner. *Funktionalanalysis*. 6. Aufl. 2007. ISBN: 978-3-540-72536-7.

## Numerische Details

Für die Implementierung wird die Funktion  $J_{dis}^{y,p}$  umgestellt, so dass insgesamt möglichst wenige Berechnungen vorzunehmen sind. Das Vorgehen dabei wird hier kurz beschrieben. Wir nummerieren die Elemente der Gitter durch, via

$$\Omega_{\#} = \{a_0, a_1, \dots, a_{|\Omega_{\#}|}\} \quad \text{und} \\ T_{\#} = \{x_0, x_1, \dots, x_{|T_{\#}|}\} .$$

Beginnend mit der  $|\Omega_{\#}| \times K$ -Matrix  $\phi$ ,

$$\phi := \begin{pmatrix} (\phi_0(a_0) + \phi_1(a_0))/2 & \cdots & (\phi_{K-1}(a_0) + \phi_K(a_0))/2 \\ \vdots & \ddots & \vdots \\ (\phi_0(a_{|\Omega_{\#}|}) + \phi_1(a_{|\Omega_{\#}|}))/2 & \cdots & (\phi_{K-1}(a_{|\Omega_{\#}|}) + \phi_K(a_{|\Omega_{\#}|}))/2 \end{pmatrix} ,$$

können wir anschließend die  $|\Omega_{\#}| \times |T_{\#}|$  Gauss-Matrix zum Zeitpunkt  $k$  definieren:

$$GM_k := \begin{pmatrix} G_{\phi(0,k)}(x_0) & \cdots & G_{\phi(0,k)}(x_{|T_{\#}|}) \\ \vdots & \ddots & \vdots \\ G_{\phi(|\Omega_{\#}|,k)}(x_0) & \cdots & G_{\phi(|\Omega_{\#}|,k)}(x_{|T_{\#}|}) \end{pmatrix} ,$$

wobei  $\phi(i, j)$  den  $i, j$ -ten Eintrag von  $\phi$  notiert. Wir definieren weiter eine  $\widehat{D} \times |\Omega_{\#}|$ -Matrix, erneut für festes  $k$ :

$$E_k := pE \cdot \begin{pmatrix} \sum_{v \in G(1,d)} \sum_{y \in P_z^{-1}(\Gamma_{(0),v})} GM_k(0, y) & \cdots & \sum_{v \in G(1,d)} \sum_{y \in P_z^{-1}(\Gamma_{(0),v})} GM_k(|\Omega_{\#}|, y) \\ \vdots & \ddots & \vdots \\ \sum_{v \in G(1,d)} \sum_{y \in P_z^{-1}(\Gamma_{(\widehat{D}),v})} GM_k(0, y) & \cdots & \sum_{v \in G(1,d)} \sum_{y \in P_z^{-1}(\Gamma_{(\widehat{D}),v})} GM_k(|\Omega_{\#}|, y) \end{pmatrix} .$$

Es gilt hier zu beachten, dass mit  $y$  das zugehörige Element  $x_y$  aus  $T_{\#}$  gemeint ist. Wir können nun für ein festes  $b \in \Omega_{\#}$  schreiben:

$$\frac{\partial}{\partial w_b} J_{dis}^{y,p}(w) = \sum_{k,i} \left[ c_{K,d,i} + E_k(i, b) - y_{k,i} \frac{p \cdot C_{K,d,i} + E_k(i, b)}{p \cdot C_{K,d,i} \|w\|_1 + (E_k \cdot w)(i)} \right] + \beta ,$$

mit  $(E_k \cdot w)(i)$  als den Eintrag der  $i$ -ten Zeile der Matrix-Vektor-Multiplikation. Mit weiteren Definitionen von Termen, die nach Erhalt der Messungen und während des Frank-Wolfe-Algorithmus konstant bleiben, erhalten wir:

$$\frac{\partial}{\partial w_b} J_{dis}^{y,p}(w) = c'(b) - \sum_{k,i} \left[ \frac{c''(k, i, b)}{p \cdot c_{K,d,i} \cdot \|w\|_1 + (E_k \cdot w)(i)} \right] + \beta , \quad \text{wobei}$$

$$c'(b) = \sum_{k,i} c_{K,d,i} + E_k(i, b) \quad \text{und} \quad c''(k, i, b) = y_{k,i} \cdot (p \cdot c_{K,d,i} + E_k(i, b)) .$$

Im letzten Schritt werden alle derart ermittelten partiellen Ableitungen in einem Vektor zusammengefasst, der dem Gradienten von  $J_{dis}^{y,p}$  entspricht. Den vorherigen Definitionen nach kann zuletzt auch  $J_{dis}^{y,p}$  umgeschrieben werden:

$$J_{dis}^{y,p}(w) \approx \sum_{k,i} \left[ c_{K,d,i} \|w\|_1 + (E_k \cdot w)(i) - y_{k,i} \log \left( p \cdot c_{K,d,i} \cdot \|w\|_1 + (E_k \cdot w)(i) \right) \right] + \beta \|w\|_1 .$$

# Eidesstattliche Erklärung

Hiermit versichere ich, *Matthias Fröhlich*, dass die vorliegende Arbeit über *PET-Bildrekonstruktion im Phasenraum* selbstständig verfasst worden ist, dass keine anderen Quellen und Hilfsmittel als die angegebenen benutzt worden sind und dass die Stellen der Arbeit, die anderen Werken – auch elektronischen Medien – dem Wortlaut oder Sinn nach entnommen wurden, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht worden sind.

Münster, 26. August 2021



Unterschrift

Ich erkläre mich mit einem Abgleich der Arbeit mit anderen Texten zwecks Auffindung von Übereinstimmungen sowie mit einer zu diesem Zweck vorzunehmenden Speicherung der Arbeit in eine Datenbank einverstanden.

Münster, 26. August 2021



Unterschrift