INAUGURAL - DISSERTATION

zur

Erlangung der Doktorwürde

der

Gesamtfakultät für Mathematik, Ingenieur- und Naturwissenschaften

der

Ruprecht - Karls - Universität

Heidelberg

vorgelegt von

M. Sc. Christian Alexander Düll

aus Heidelberg

Tag der mündlichen Prüfung

____

Generalising nonlinear population models

Radon measures, Polish spaces and the flat norm

Supervisor

Prof. Dr. Anna Marciniak-Czochra

Abstract

This thesis deals with Radon measures and how they can be used to extend nonlinear structured population models from the classical Euclidean state spaces to abstract Polish metric spaces. To this end, we first investigate the functional analytic properties of the space of Radon measures under the flat norm and show that in some sense the latter generalises the well-known Wasserstein distance W_1 from the conservative to the unbalanced case. We then apply variational inequality theory to derive an explicit formula for the flat distance between a linear combination of Dirac deltas and a fixed Dirac measure.

The second part of the thesis deals with structured population models in measures. As a start, we consider the Euclidean case and prove well-posedness of the linear model, first without and then with state-independent influx, using Duhamel's principle. Subsequently, the nonlinear model is solved via a reduction to a linear model with frozen measure arguments.
The gathered insights in the $\mathbb{R}^d$ case enable us to abstract the necessary concepts so that we can transfer the model formulation to general Polish metric spaces. However, in the absence of a vector space structure, we cannot rely on a governing differential equation, and introduce a suitable implicit integral representation instead which serves both as model and notion of solution. A major step towards well-posedness of the involved Bochner integrals is given by the structure of the space of measures and its favorable properties under the flat norm.

We conclude with various outlooks on the applicability of our theory. At first, several ideas for promising models in measures that can fully exploit the abstract Polish state spaces are sketched. Afterwards, we show how to incorporate measure differential equations and - with a minimal adaptation- also the class of coagulation-fragmentation models into our framework. Finally, it is proven that our flat norm is closely related to a transport type distance developed by Fournier and Perthame to study the asymptotics of nonexpanding transport processes.

Zusammenfassung

Diese Arbeit befasst sich mit der Erweiterung nichtlinearer strukturierter Bevölkerungsmodelle von den klassischen euklidischen Zustandsräumen auf polnische metrische Räume. Zu diesem Zweck nutzen wir eine Formulierung in Maßen und untersuchen die Eigenschaften des Raums der Radon Maße unter der flachen Norm. Weiterhin zeigen wir, dass die flache Norm im gewissen Sinne die bekannte Wasserstein-Distanz W_1 vom konservativen zum unausgewogenen Fall verallgemeinert. Anschließend wenden wir die Theorie der Variationsungleichungen an, um eine explizite Formel für den flachen Abstand zwischen einer Linearkombination von Dirac-Deltas und einem festen Dirac-Maß herzuleiten.

Der zweite Teil der Arbeit beschäftigt sich mit strukturierten Populationsmodellen in Maßen. Zunächst betrachten wir den euklidischen Fall und beweisen die Wohldefiniertheit des linearen Modells, zunächst ohne und dann, unter Verwendung des Duhamelschen Prinzips, mit zustandsunabhängigem Zufluss. Anschließend wird das nichtlineare Modell durch eine Reduktion auf ein lineares Modell mit eingefrorenen Maßargumenten gelöst. Die gewonnenen Erkenntnisse im $\mathbb{R}^d$-Fall ermöglichen es uns dann, die notwendigen Konzepte soweit zu abstrahieren, dass wir die Modellformulierung auf allgemeine polnische metrische Räume übertragen können. In Ermangelung einer Vektorraumstruktur können wir uns jedoch nicht auf eine Differentialgleichung berufen und führen stattdessen eine geeignete implizite Integraldarstellung ein, die sowohl als Modell als auch als Lösungsbegriff dient.

Abschließend geben wir verschiedene Ausblicke auf die Anwendbarkeit unserer Theorie. Dafür werden zunächst einige Ideen für Modelle in Maßen skizziert, die die abstrakten polnischen Zustandsräume voll ausnutzen können. Danach zeigen wir, wie Maßdifferentialgleichungen und - mit einer minimalen Anpassung - auch die Klasse der Koagulations-Fragmentierungsmodelle in unserem Ansatz mit einbezogen werden können. Zum Schluss untersuchen wir die enge Verwandschaft unserer flachen Norm zu einer Transportdistanz, die von Fournier und Perthame für das Studium der Asymptotik nicht expandierender Transportprozesse entwickelt wurde.

Acknowledgements

First of all, I want to express my deepest gratitude to my supervisor Prof. Dr. Anna Marciniak-Czochra who, without knowing me, immediately accepted me into her working group. I still greatly appreciate the opportunity she gave me as a young Master's graduate to co-author a scientific textbook. This honor is not given to many doctoral students and the decision influenced my entire doctoral career from the fruitful collaborations with Warsaw to my teaching at university all the way to the topics of this thesis. Even in more chaotic times, such as the Corona pandemic or the birth of my children, she always supported me and allowed me to complete my projects and duties in my own way and at my pace.
I would also like to thank the countless people I have had the pleasure of meeting during my time at the faculty, especially Chris, Maria, Filip, Frits, Finn, Denis, Carolin and Theo. It was the shared conversations, Christmas parties, lunches and (mathematical) discussions that made my time here come alive. Outside of Heidelberg University, I especially would like to thank Kuba, who was always open to my problems and who more than once pointed me in the right direction.

Last but not least, I want to thank my family for their constant support, not only during my doctoral studies but also during my entire time in Heidelberg. Because of you, I never lost the ground under my feet and also got some much needed distraction when my mathematical problems seemed unsolvable. Special thanks goes to my wife for being my soul mate and strong partner in all situations. Without you, I might have never knocked on Anna's office door in the first place. I hope that we will have many more adventures together.

Notation

Basics

$\mathbb{N}$, $\mathbb{N}_0$	Set of all natural numbers; $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$
$\mathbb{R}$, $\mathbb{R}^+$, $\overline{\mathbb{R}}$	Set of all real numbers; $\mathbb{R}^+ = [0, \infty)$; $\overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$
$\bigotimes_{i=1}^N A_i$	Cartesian product of the sets A_i, $i = 1, ..., N$
$\mathcal{B}(S)$	Borel sigma algebra of the set S
$\operatorname{diam}(X)$	Diameter of the metric space (X, d); $\operatorname{diam}(X) = \sup_{x,y \in S} d(x, y)$
$B_r(x)$	Open ball of radius r around x; $B_r(x) := \{y \in S \mid d(x, y) < r\}$
$B_r(A)$	Open r-neighbourhood of A; $B_r(A) := \{y \in S \mid d(y, A) < r\}$
$\operatorname{lin}(A)$	Linear span of A; $\operatorname{lin}(A) = \{\sum_{k=1}^n \alpha_k x_k \mid n \in \mathbb{N}, \alpha_k \in \mathbb{R}, x_k \in A\}$
$\operatorname{supp}(f)$	Support of the function f, $\operatorname{supp}(f) = \overline{\{x \in S \mid f(x) \neq 0\}}$
$\operatorname{supp}(\mu)$	Support of the measure μ; $\operatorname{supp}(\mu) = \overline{\{A \in \mathcal{B} \mid \mu(A) \neq 0\}}$

Special Functions

$\mathbf{1}$	Constant 1 function
χ_A	Characteristic function of the set A; $\chi_A(x) = \begin{cases} 1, & x \in A \\ 0, & \text{else} \end{cases}$
$f^\pm$	Positive resp. negative part of the function f; $f^\pm := \max\{\pm f, 0\}$
b	Vector field of the transport term
c	Linear growth term
η	Mutation kernel
N	State-independent influx
X_b	Flow of the vector field b; solution to the ODE (3.2.1)
$\varphi_{\psi,t}$	Solution to the dual problem (3.3.1)

Special Measures

λ^d	Lebesgue measure on $\mathbb{R}^d$
δ_x	Dirac measure located at x, $\delta_x(A) = \begin{cases} 1, & x \in A \\ 0, & \text{else} \end{cases}$
$D_{\lambda^d}\mu$	(Radon-Nikodým) Density of μ with respect to λ^d
$\mu_\bullet$	Short for $\{\mu_t\}_{t \in [0,T]}$
$T_\# \mu$	Push-forward of μ under the map T; $T_\#\mu(\cdot) = \mu(T^{-1}(\cdot))$

(Semi-) Norms and Metrics

$\lvert x \rvert$	Standard Euclidean norm on $\mathbb{R}$ or $\mathbb{R}^d$
$\lvert f\rvert_{\mathbf{Lip}}, \lvert f\rvert_{\mathbf{Lip}_d}$	Lipschitz constant of $f:(S,d)\to\mathbb{R}$; $\lvert f\rvert_{\mathbf{Lip}} := \sup_{x\neq y}\frac{\lvert f(x)-f(y)\rvert}{d(x,y)}$
$\lVert f\rVert_\infty$	Supremum norm of $f:(S,d)\to\mathbb{R}$; $\lVert f\rVert_\infty := \sup_{x\in S}\lvert f(x)\rvert$
$\lVert f\rVert_{BL}$	Bounded Lipschitz norm of $f:(S,d)\to\mathbb{R}$; $\lVert f\rVert_{BL} := \max\{\lVert f\rVert_\infty, \lvert f\rvert_{\mathbf{Lip}}\}$
$\lVert f\rVert_\lambda$	Bielecki norm; $\lVert f\rVert_\lambda = \sup_{(\tau,x)\in[0,T]\times\mathbb{R}^d} e^{-\lambda(T-\tau)}\lvert f(\tau,x)\rvert$
$\lVert f\rVert_{\lambda,g}$	Generalised Bielecki norm; see (4.2.7)
$\lvert\mu\rvert(A)$	Variation of μ; see Definition 2.1.2
$\lVert\mu\rVert_{TV}$	Total variation norm; $\lVert\mu\rVert_{TV} := \lvert\mu\rvert(S)$
$\lVert\mu\rVert_{BL^*}, \rho_F$	Flat norm and corresponding flat metric; $\lVert\mu\rVert_{BL^*} := \sup_{\lVert\psi\rVert_{BL}\leq 1}\int_S \psi\,\mathrm{d}\mu$
$W_1(\mu,\nu)$	Wasserstein-1 distance; $W_1(\mu,\nu) := \sup_{\lvert\psi\rvert_{\mathbf{Lip}\leq 1}}\int_S \psi\,\mathrm{d}(\mu-\nu)$
$\lVert\eta\rVert_{L^1_T}$	L^1 norm of η; $\lVert\eta\rVert_{L^1_T(BL(S;\mathcal{M}^+))} := \int_0^T \lVert\eta(t,\cdot)\rVert_{BL(S;\mathcal{M}^+)}\,\mathrm{d}t$

Spaces

$C^k(X;Y)$, $C^k(X)$	Space of k-times continuously differentiable functions $X\to Y$; $C^k(X) := C^k(X;\mathbb{R})$
C^k_c; C^k_b	C^k functions which are compactly supported resp. bounded
$C_0(X)$	Space of continuous functions $X\to\mathbb{R}$ vanishing at infinity
$L^1(\mu)$	Space of functions which are integrable with respect to μ
$BL(X;Y)$	Space of bounded Lipschitz functions $X\to Y$
X^*	Dual space of X; $X^* := \{T: X\to\mathbb{R} \mid T \text{ linear and bounded}\}$
$BL(S)^*_+$	Space of positive linear functionals on $BL(S)$; $BL(S)^*_+ := \{T\in BL(S)^* \mid T(\psi)\geq 0\,\forall\psi\in BL(S), \psi\geq 0\}$
$\mathcal{M}(S)$	Space of finite and signed Borel measures on S
$\mathcal{P}(S)$	Space of probability measures on S
$\mathcal{M}^+(S)$	Cone of nonnegative finite Borel measures on S
$\mathcal{M}^+_1(S)$	$\mathcal{M}^+(S)$ measures with integrable first moment; $\mathcal{M}^+_1(S) := \left\{\mu\in\mathcal{M}^+(S) \mid \int_S d(x,x_0)\,\mathrm{d}\mu < \infty\right\}$
E	Closure of $\mathcal{M}(S)$ under the flat norm; $E := \overline{\mathcal{M}(S)}^{\lVert\cdot\rVert_{BL^*(S)}}$

Contents

Abstract iii

Zusammenfassung v

Notation ix

1 Introduction 1

2 The space of measures 11

2.1 Basic measure theory . . . 11
2.2 The flat norm . . . 15
2.3 The space of measures under the flat norm . . . 17
2.4 The flat norm as a distance for unbalanced optimal transport . . . 23
2.5 Computing the flat metric . . . 39

3 PDE models on the Euclidean space 44

3.1 From a PDE model to a weak formulation . . . 44
3.2 The flow of a vector field . . . 50
3.3 The case without state-independent influx . . . 53
3.4 Adding state-independent influx to the model . . . 83
3.5 The nonlinear case . . . 91
3.6 An extension of the linear case- uniform tightness . . . 100

4 Models on Polish metric spaces 104

4.1 Two-parameter families of Lipschitz maps . . . 104
4.2 The generalised linear structured population model . . . 105
4.3 The generalised nonlinear model . . . 117
4.4 Model equivalence in the Euclidean case . . . 122

5 Future perspectives and connection to other publications 136

5.1 Model examples motivated by applications . . . 136
5.2 Connection to measure differential equations . . . 138

5.3 A structured coagulation-fragmentation equation 141
5.4 A nonexpanding transport distance 146

6 Conclusion and outlook **153**

Related publications **155**

Bibliography **156**

1 Introduction

Mathematical methods and in particular differential equations have long been an integral and successful component in the modelling and analysis of physical, chemical and biological systems. They allow an abstract view of the underlying mechanisms and are often able to identify the essential components of the involved processes. Conversely, mathematical methods must constantly adapt to new research topics and experiments of the applied sciences. From time to time these even provide impulses for the development of new mathematical concepts when the existing approaches are insufficient.

Specifically, in this thesis we consider problems that address the dynamics of heterogeneous populations such as cells, animals and plants which can be structured with respect to an underlying structuring or state variable [40], so called *structured population models*. As it stands, these kind of models are used in a broad spectrum of applications ranging from cell biology [157, 160], immunology [110], epidemiology [30, 41] and stage of cell differentiation [42, 72, 78, 168] to demography [20, 57, 116], ecology [86, 116] and environmental economics [8].

Typically, structured population models apply ordinary differential equations (ODEs), stochastic differential equations (SDEs) or partial differential equations (PDEs) to describe the evolution of the population and use the structuring variable to incorporate information on physical or biological properties. A common distinction is drawn between *intrinsic state variables*, which often describe *physiological* characteristics of the individuals in the population including age [32, 68], size [5, 42, 47, 82] or stage of infection [108, 162], and *extrinsic variables*, such as spatial distribution induced by heterogeneity of the environment [6, 16, 74, 91].

For simplicity, in many cases the structured variable is assumed to be discrete, which leads to compartmental models describing birth-death processes with transitions between the different discrete states. Despite the limiting assumption of discrete states, compartmental models are still powerful enough to capture the essential aspects of many processes [36, 139, 144]. For example, in Refs. [115, 142, 163] compartmental models significantly contributed to the quantification of stem cell traits such as proliferation, self-renewal or quiescence in developmental and regeneration processes

and in cancer, see also [105, 145]. Furthermore, they revealed novel aspects of the role of systemic and microenvironmental feedback loops and highlighted heterogeneity [143], clonal evolution [141] and cell-cell interactions [93, 134, 146].
However, due to the recent advances in high dimensional data acquisition, like single cell sequencing methods, biologists are now able to collect data on an unprecedented scale which occasionally challenges established knowledge or puts it in a different context. For example, it was long assumed that populations of cells in the same cell state were homogeneous, but in view of the new data resolution they turned out to be remarkably heterogeneous. Even more, ordering transcriptomes along a one-dimensional "pseudotime" curve using statistical data analysis as well as in silico pathway analysis and functional assays, showed that certain functionally different cell states exhibit a fair degree transcriptional similarity, suggesting the existence of some continuous cell state transitions [61, 84].
Classically, continuous transitions are either modelled by transport-type PDEs or using delay differential equations (DDEs). The latter apply ODEs together with delays to express, for example, duration of immunity [11], maturation time [5, 71, 83, 165] or effect of predation [127]. Due to the delayed time dependence DDE systems are inherently infinite dimensional. Specifially for structured population models, the corresponding delay formulation has enabled new techniques for the analysis of linearised stability and Hopf bifurcations [20, 38, 39].
That said, the analysis of single-cell data also revealed several aspects that cannot be directly addressed in a classical PDE or DDE setting, such as the coexistence of continuous and discrete (jump) transitions [78], loops or multiple structures. Consequently, linking the single cell data to mechanistic mathematical modelling of the underlying processes requires a new formulation of the models admitting the envisioned complexity of the cell state space [36].
Hence, in this thesis, we focus on a promising approach that has undergone a strong development in recent years, namely formulating PDE models in the space of Radon measures [1, 23, 55, 56, 81, 150, 151, 152]. Measures have the advantage that they can not only represent discrete and continuous structures but also mixtures of both [2], thus allowing for irregular behaviour such as jumps or concentration effects. Furthermore, the underlying state spaces for measures require very little structure which makes them ideal candidates for formulating models on non-Euclidean spaces, including graphs, networks or manifolds. For a convenient usage of measures in the PDE models, though, a unified well-posedness theory for such problems was missing and so existence and uniqueness of solutions had to be shown for each new problem

individually [3, 54, 65, 66, 79, 80]. This changed with the publication of the book Ref. [44] which extended a broad class of structured population models of the form

$$\begin{cases} \partial_t \mu_t + \nabla_x \cdot (b(t,x,\mu_t)\mu_t) &= c(t,x,\mu_t)\mu_t + \int_{\mathbb{R}^d} \eta(t,x,\mu_t)(y)\,\mathrm{d}\mu_t(y) + N(t,\mu_t), \\ \mu_0 &= \nu, \end{cases} \tag{1.1.1}$$

to generalised models on proper metric spaces and established a suitable well-posedness theory. We refer to Remark 2.3.5 for a definition of properness and to the beginning of Chapter 3 for a thorough description of the model functions. Since metric spaces S generally lack a linear structure, the concept of derivatives is unavailable and thus the PDE (1.1.1) does not make sense on the state space S. So instead of characterising the transport process using a differential equation, we establish an explicit description using a push-forward operator. This results in an implicit integral representation which is motivated by the method of characteristics in the Euclidean case,

$$\begin{aligned} \mu_t &= X(t,0,\cdot,\mu_\bullet)_\# \left(\mu_0(\cdot) e^{\int_0^t c(s,\,X(s,0,\cdot,\mu_\bullet),\,\mu_s)\,\mathrm{d}s}\right) \\ &+ \int_0^t X(t,\tau,\cdot,\mu_\bullet)_\# \left(\int_S [\eta(\tau,y,\mu_\tau)(\cdot)]\,\mathrm{d}\mu_\tau(y) e^{\int_\tau^t c(s,\,X(s,\tau,\cdot,\mu_\bullet),\,\mu_s)\,\mathrm{d}s}\right)\mathrm{d}\tau \\ &+ \int_0^t X(t,\tau,\cdot,\mu_\bullet)_\# \left(N(\tau,\mu_\tau)(\cdot) e^{\int_\tau^t c(s,\,X(s,\tau,\cdot,\mu_\bullet),\,\mu_s)\,\mathrm{d}s}\right)\mathrm{d}\tau. \end{aligned} \tag{1.1.2}$$

On the abstract state space, the implicit representation serves both as the model and as the notion of solution, see Chapter 4 for a detailed description. However, the assumption that the underlying space is proper turns out to be quite strong. Suppose for a moment that the space is actually normed. In this case, basic functional analysis yields that S is necessarily finite dimensional [131, 1.22]. In order to avoid this limitation, the subsequent paper Ref. [46] showed that separable and complete metric spaces, so-called Polish metric spaces, are already sufficient to guarantee well-posedness of the generalised model (1.1.2) and thus allowing for infinite-dimensional state spaces. The results obtained in Refs. [44, 46] serve as the basis for this thesis and as a starting point for further studies which extend or complement the existing theory. A detailed description of the results is given in the Section *Scientific Contribution*.

From the modelling perspective, a unified well-posedness theory for models defined on a rich class of state spaces provides some interesting possibilities. On the one hand, there are already several approaches to infer governing PDEs from data with the help of deep neural networks [103, 104, 140]. In this setting, our framework could form the basis for learning discrete as well as continuous processes simultaneously without having prior knowledge or limiting assumptions regarding the structure of the process. At the same time, our theory also allows us to investigate the influence of the state space on the resulting models. For example, recently two similar cell state models have been compared, which differ only in being formulated on an underlying continuous or graph-based state space [29]. However, due the lack of a uniform setting both alternative models had to be studied separately. In the future, such questions could hopefully be addressed with the help of inverse problems using mechanistic models of biological processes and associated data. For the analysis of single cell data, the question of the state space structure is essential, and in this context, model-based approaches can be a valuable instrument alongside the topology-based [112, 125, 128], geometric [95] or statistical techniques currently used for data analysis.

From a mathematical viewpoint, our framework is designed to generalise transport-type PDEs to models beyond the classical $\mathbb{R}^d$ setting and is thus not limited to models of mathematical biology, but can also be used in the fields of pedestrian flows, crowd dynamics [13, 14, 15] or fluid motion [90]. There are currently several approaches considering pedestrian flows on (road) networks [22, 88] and ideal fluids have already been studied in a noneuclidean setting, at least on spheres [137]. So the possibility of allowing for more general state spaces, such as the torus, could enable a plethora of new interesting research topics.

Furthermore, the model formulation in measures on Polish metric spaces creates a connection to concepts and methods developed in stochastic modelling, in particular to the theory of concentrating Feller operators. These provide encouraging results for the asymptotic analysis of measure solutions [99, 147, 148, 153]. For example, the asymptotics of transport-type equations can be linked to ergodic properties of Markov processes [85]. Further asymptotic results of certain measure valued solutions can be found in Ref. [58, 67].

In addition to the generalised modelling framework, the implicit integral representation (1.1.2) reveals an intriguing relationship to the theory of optimal transport (OT) problems. For simplicity, consider the conservative case with $c, \eta, N \equiv 0$ in which the initial measure μ_0 is just moved according to the map $X(t, \cdot, 0)$. In this scenario, formula (1.1.2) reduces to $\mu_t = X(t, \cdot, 0)_{\#}\mu_0$, i.e. the solution μ_t solves an OT problem with transport map $X(t, \cdot, 0)$ (see [159, p.16-17]). Differently speaking, from an OT perspective formula (1.1.2) provides a substantial extension to the simple (linear) transport case by introducing several additional effects that can influence the movement of the optimal flow. For instance, this allows growth processes to be incorporated in order to consider unbalanced OT problems, such as those used to determine cell fate trajectories [133, 167]. Conversely, if at some point methods for parameter estimation are available for our framework, it raises the hope that for explicit processes appearing in applications it will be possible to distinguish which effects are responsible for which part of the flow dynamics.

A similar, but slightly different approach to the one discussed in this thesis is given by the theory of *measure differential equations* (MDEs) [21, 24, 45, 119, 123] which use a measure-valued transport velocity to describe for example pedestrian flows in the Euclidean space. The transport velocity is prescribed by a so-called *measure vector field* (MVF), a nonnegative measure on the tangent space of $\mathbb{R}^d$ that assigns a distribution of possible transport velocities to each point $x \in \mathbb{R}^d$. So in contrast to the usual transport models, MVFs allow for more general directions in each point. This can be particularly interesting for modelling RNA velocities, which were previously limited to one main direction of movement [98, 106]. Newer methods based on deep neural networks instead calculate empirical posterior distributions of RNA velocities and thus achieve better and more stable results [69]. However, as we show in Section 5.2 most MVFs can be represented by a particular choice of the flow b, and thus MDEs are merely a special case of our framework.

Scientific Contribution

With this dissertation I extend and complement the results that have been developed in Refs. [44, 46]. The first part of this thesis is thus devoted to the space of Radon measures $\mathcal{M}^+$ under the flat norm and its general functional analytical properties. To better understand the behaviour of the flat norm, Piccoli's alternative characterisation [121], which connects the flat metric with the total variation norm and the Wasserstein distance W_1, is extended from $\mathbb{R}^d$ to general proper spaces. This then enables to derive a closed formula for the flat distance between measures of the form

$$\rho_F\left(c\delta_{x_0}, \sum_{i=1}^{N} b_i\delta_{x_i}\right)$$

using variational inequality theory [70]. A closed formula for the distance between any two linear combinations of Dirac measures is unfortunately out of reach due to the lack of a corresponding result for the Wasserstein distance. This part of the thesis then concludes with a detailed summary of the preprint Ref. [135] which uses a neural network (NN) to compute the flat norm between two arbitrary measures. As a starting point, we use the results of Anil, Lucas and Grosse, which apply a NN to calculate the Wasserstein distance W_1 between two measures [9]. Due to the similarity of the Kantorovich-Rubinstein duality with the definition of the flat metric, we are in the position to preserve the structure of the original network and only modify the corresponding loss function.

The second part of the thesis addresses well-posedness theory of structured population models in measures. In this context, I first focus on models in $\mathbb{R}^d$ of the form (1.1.1), which have not yet been investigated in this generality. In particular, the state independent influx N ensures that the associated differential operator is not linear in the solution μ_t, and thus passing to the corresponding adjoint operator is not possible. To overcome this problem, I first consider the linear version of the model without the disturbing state-independent influx. In this case, the dual problem can be solved with the method of characteristics and a fixed point argument, providing an implicit representation formula for the dual solution $\varphi_{\psi,t}$, see Proposition 3.3.1. However, in contrast to the model analysed in Chapter 2 of [44], in this thesis the model functions are assumed to be time-dependent, and thus the construction of a measure solution to the primal problem from $\varphi_{\psi,t}$ requires additional effort. In particular, Fréchet differentiability of the map $t \mapsto \varphi_{\psi,t}$ is needed to show that the measure solution μ_t actually satisfies the weak formulation. The necessary regularity is established in Proposition 3.3.16 using the Implicit Function Theorem in Banach spaces. Subsequently, well-posedness as well as Lipschitz continuity results are shown.

In a second step, the state-independent influx is added to the linear model. Using Duhamel's principle, a solution to the modified problem is constructed and the corresponding regularity assumptions are transferred. At this point, one can also prove uniqueness of the solution to the full linear problem by resorting to a suitably chosen dual problem. In analogy to the procedure in Ref. [44], the nonlinear problem is then solved via a reduction to a linear problem. To this end, the time interval is divided into an equidistant grid and the nonlinearities are frozen, resulting in a linear problem. Using the established estimates of the linear case, one can show that the approximative solutions form a Cauchy sequence and thus converge to a solution of the nonlinear problem. However, this approach only provides existence and continuous dependence of the solution, whereas uniqueness will only follow from the methods in Chapter 4.

The treatment of the Euclidean case concludes with the observation that the solutions constructed here are bounded on the considered time interval $[0, T]$ for any T, and therefore the measure solution μ_t can be continued for all times. However, in this situation the mass may escape to infinity. To prevent this undesirable behaviour, the whole family of solutions $\{\mu_t \mid t \in [0, \infty)\}$ has to be tight which in the linear case can be achieved by the Theorem of Prokhorov together with upgraded regularity assumptions to establish bounds uniform in T. The tight extension of the measure solution for all times is the first step for analysing the asymptotics of measure solutions using the methods developed for concentrating Feller operators, see [99, 147, 148]. Though, this will only be the subject of future studies.

The next part of the thesis focuses on the extension of models in measures to Polish metric spaces, where the presented results led to the publication Ref. [46]. Metric spaces generally lack a linear structure so that one cannot rely on the method of characteristics for constructing solutions. Consequently, there is also no need to ensure sufficient regularity of the dual solution and thus the model functions manage with less regularity assumptions in time, compared to Chapter 3. Instead of the PDE (1.1.1), the implicit integral representation (1.1.2) is used both as a model and notion of solution. In the formula, X denotes the generalisation of the flow of the vector field X_b which in turn is the solution of the ODE

$$\partial_t X_b(t; \tau, x, \mu_\bullet) = b(t, X_b(t; \tau, x, \mu_\bullet), \mu_t) \qquad X_b(\tau; \tau, x, \mu_\bullet) = x.$$

The assumptions that were placed on X in Ref. [46] were strongly inspired by the properties of the flow X_b. In this thesis, however, it turned out that Lipschitz continuity of X provides sufficient regularity for all the proofs, so that the former assumption of a bi-Lipschitz homeomorphism could be relaxed significantly. Using a fixed point argument and the implicit integral representation (1.1.2), unique generalised solutions can be constructed to both the linear and the nonlinear model. Furthermore, it is shown that the new framework is indeed a natural extension of the PDE model since the generalised and the measure solution to the PDE (1.1.1) coincide in the Euclidean case.

The last part of the thesis deals with possible applications of the presented theory. After sketching a few potential models, I demonstrate that the concept of measure

differential equations (MDEs) introduced by Piccoli in [119, 123], can already be reproduced by the theory presented in this thesis. More precisely, central element of the MDEs is the measure-valued transport velocity, called measure vector field (MVF) which can be seen as a velocity distribution at each location. Using Disintegration Theorem together with a suitable choice of the flow b can mimic the modelling effect of a MVF.
Another important class of models, the coagulation-fragmentation models, cannot be realised directly by the framework discussed in Chapter 3, and so Lyons, Ackleh an Saintier had to establish a separate- but quite elegant- well-posedness theory [2]. One of the key model functions in this type of equation is the coagulation kernel, which describes how particles of size x coagulate with particles of size y. This simultaneous x and y dependence translates to the test functions of the weak formulation which is not intended in the framework described above. However, I modified the weak formulation by adding a slightly adapted mutation kernel η which can account for this kind of dependencies. Fortunately, the modification does not significantly change the estimates of the original framework, so that the coagulation-fragmentation models can now be treated without any problems. Additionally, the established results on the continuous dependency of solutions by can also be applied directly to the model treated in [2].
Finally, I consider the transport distance $\mathcal{T}_\rho$ developed by Fournier and Perthame for the study of nonexpanding transport equations. In the cases in which $\mathcal{T}_\rho$ is actually a metric, it is a modified version of the Wasserstein distance W_1. I was able to show that $\mathcal{T}_\rho$ is strongly equivalent to a flat metric with an adapted distance on the underlying state space. Furthermore, the classical flat metric and the transport distance are both complete and metricise the narrow topology and are thus at least topologically equivalent. This is another important intermediate step for asymptotic analysis of measure solutions.

Outline of the thesis

In Chapter 2 we shortly introduce the most important concepts from measure theory that are necessary for the remainder of this thesis and prove basic properties of the space measures under the flat norm. Furthermore, we explore the connection of the flat norm to optimal transport theory by generalising the alternative characterisation of Piccoli to proper metric spaces. Afterwards, a closed formula for the flat distance between a Dirac delta and a linear combination of Diracs is proven using variational inequality theory. We conclude this chapter by a summary of preprint which exploits machine learning to compute the flat distance between arbitrary distributions when no analytical ground truth is available.

The heart of this thesis deals with structured population models in measures, first on $\mathbb{R}^d$ (Chapter 3), where we can work with a PDE formulation and then on arbitrary Polish metric spaces (Chapter 4). To solve the Euclidean case, we first consider a linear version of the PDE without state independent influx N in Section 3.3, so that the differential operator is linear in the solution and we can thus pass to the corresponding adjoint operator, i.e. the dual problem. The method of characteristics provides an implicit integral representation, and hence a unique solution to the dual problem can be constructed with a fixed point argument. We then use the dual problem and its solution to solve the primal problem. In Section 3.4, the state-independent influx N is added and a corresponding solution is constructed using Duhamel's principle. Furthermore, uniqueness for the linear model is established. In order to solve the nonlinear case in Section 3.5, we first divide the time interval into subintervals and freeze the nonlinearities which leads to a linear model. Using the already established results from the linear case, it can be shown that the approximative solutions form a Cauchy sequence and thus converge to a solution of the nonlinear problem by completeness. Note, however, that at this point uniqueness for the nonlinear case is out of reach, but will only result from the methods developed in Chapter 4. At the end of Chaper 3, we prove tightness of solutions to the linear model for all times. To this end, we have to upgrade the assumptions to get estimates which are uniform in T.

In Chapter 4, we utilize the insights from the previous chapters to abstract the relevant concepts as far as possible and are thus able to formulate a generalised version of the structured population model for Polish metric spaces which does not rely on differential equations. Instead of a PDE model, we use an implicit integral representation. Here, the convenient functional analytic properties of the space of Radon measures $\mathcal{M}^+$ under the flat norm guarantee that the occuring Bochner integrals are well-defined. Existence and uniqueness are proven by a fixed point argument, first in the linear (Section 4.2) and then in the nonlinear case (Section 4.3). The chapter is concluded with Section 4.4, where it is shown that both modelling approaches coincide in the $\mathbb{R}^d$ case and, while simultaneously, establishing the missing uniqueness for the PDE model of Chapter 3.

In Chapter 5 we place the new framework in the context of current research. To this end, we sketch several model applications which make full use of the abstract state spaces. Additionally, we show how we can include the concept of measure differential equations. By a minimal modification of the mutation kernel η, we can even incorporate the class of coagulation-fragmentation models into our framework, see Section 5.3. Finally, we show that our flat norm is strongly related to a transport distance introduced by Fournier and Perthame in [65] for nonexpansive transport equations.

2 The space of measures

In this chapter we introduce the main concepts of measure theory that are essential for the rest of this thesis. Apart from basic measure theoretic principles, we study the functional analytic properties of the space measures under the flat norm and examine the behaviour of the corresponding flat metric.

As always, continuous functions from a space X to a space Y are denoted by $C^0(X;Y)$. The subset of bounded continuous functions is indicated with a subscript b, i.e. $C_b^0(X;Y)$, and similarly compactly supported functions are denoted by $C_c^0(X;Y)$. If $Y = \mathbb{R}$, we omit the second argument and write $C^0(X)$. Unless stated otherwise, we will consider a metric space (S,d) which is assumed to be **Polish**, i.e. separable and complete.

2.1 Basic measure theory

In this thesis we will work with signed Borel measures on (S,d) and we refer to [62, Chapter 3.1] for basic definitions. In view of Hahn-Jordan Decomposition Theorem, any signed measure μ has a unique representation $\mu = \mu^+ - \mu^-$, where μ^+, μ^- are nonnegative measures which are mutually singular. If both $\mu^+(S), \mu^-(S) < \infty$, μ is said to be **finite** and the **space of all finite Borel measures** on (S,d) is denoted by $\mathcal{M}(S)$. According to [94, Theorem 13.6] those measures are **Radon**. The corresponding **cone of nonnegative measures** is defined as

$$\mathcal{M}^+(S) = \{\mu \in \mathcal{M}(S) \mid \mu \geq 0\},$$

where the partial ordering "$\leq$" on $\mathcal{M}(S)$ is given setwise, i.e. $\mu \leq \nu$ iff $\mu(A) \leq \nu(A)$ for all $A \in \mathcal{B}(S)$.

In some cases, we have to consider the subspace of **nonnegative measures with integrable first moment** which is given by

$$\mathcal{M}_1^+(S) := \left\{ \mu \in \mathcal{M}^+(S) \,\middle|\, \int_S d(x,x_0)\,\mathrm{d}\mu < \infty \right\},$$

and $x_0 \in S$ being an arbitray point. It is easy to see that $\mathcal{M}_1^+(S)$ does not depend on the choice of x_0. As usual, for any $\mu \in \mathcal{M}(S)$ the space $L^1(\mu)$ denotes the set of functions $S \to \mathbb{R}$ which are **integrable with respect to** μ.

When treating measures in $\mathcal{M}(S)$, we will use the following notion of convergence.

Definition 2.1.1. *A sequence of measures $(\mu_n)_{n\in\mathbb{N}} \subset \mathcal{M}(S)$* ***converges narrowly*** *to a measure $\mu \in \mathcal{M}(S)$, if the sequence converges in duality with all continuous and bounded functions, i.e.*

$$\lim_{n\to\infty} \int_S \psi(x)\,\mathrm{d}\mu_n(x) = \int_S \psi(x)\,\mathrm{d}\mu(x) \qquad \forall \psi \in C_b(S).$$

A map $\mu_\bullet : [0,T] \to \mathcal{M}(S)$, $t \mapsto \mu_t$ is called ***narrowly continuous*** *if $\mu_\bullet$ is continuous with respect to narrow convergence, i.e.*

$$\lim_{s\to t} \int_S \psi(x)\,\mathrm{d}\mu_s(x) = \int_S \psi(x)\,\mathrm{d}\mu_t(x) \qquad \forall \psi \in C_b(S) \text{ and } \forall t \in [0,T].$$

Definition 2.1.2. *(i) For a measure $\mu \in \mathcal{M}(S)$ with corresponding Hahn-Jordan decomposition μ^+, μ^- the* ***variation*** *$|\mu|$ of μ is defined by*

$$|\mu|(A) := \mu^+(A) + \mu^-(A) \qquad \forall A \in \mathcal{B}(S).$$

(ii) A set $\mathcal{N} \subseteq \mathcal{M}(S)$ is called ***tight*** *if for each $\varepsilon > 0$ there exists a compact set $K \subseteq S$ with*

$$|\mu|(S \setminus K) < \varepsilon \quad \forall \mu \in \mathcal{N},$$

In particular, a sequence $(\mu_n)_{n\in\mathbb{N}} \subset \mathcal{M}(S)$ is ***tight****, if $\mathcal{N} := \{\mu_n \mid n \in \mathbb{N}\}$ is tight and a single measure μ is* ***tight****, if $\{\mu\}$ is tight.*

The famous Theorem of Prokhorov provides an important connection between tightness and sequential compactness in the topology induced by narrow convergence. We omit the proof here and instead refer to [44, 1.55] where we presented the elegant proofs of [35, Theorem 2.3] and [96, Theorem 8.9].

Theorem 2.1.3. *Let (S,d) be Polish and $\mathcal{F} \subset \mathcal{M}^+(S)$ with $\sup_{\mu\in\mathcal{F}} \mu(S) < \infty$. Then the following statements are equivalent.*

i) $\mathcal{F}$ is tight.

ii) The closure of $\mathcal{F}$ with respect to narrow convergence is sequentially compact.

Corollary 2.1.4. *Let (S,d) be Polish and $(\mu_n)_{n\in\mathbb{N}} \subset \mathcal{M}^+(\mathbb{R}^d)$. If $\mu_n \to \mu \in \mathcal{M}^+(\mathbb{R}^d)$ narrowly, then $\{\mu_n \mid n \in \mathbb{N}\}$ is tight.*

Proof. Narrow convergence implies that

$$\mu_n(S) = \int_S 1 \,\mathrm{d}\mu_n \to \int_S 1 \,\mathrm{d}\mu = \mu(S) < \infty,$$

and in particular $\sup_{n\in\mathbb{N}} \mu_n(S) < \infty$. By construction the set $\{\mu_n \mid n \in \mathbb{N}\}$ is relatively sequentially compact with respect to narrow convergence so that Theorem 2.1.3 yields the claim. □

Next, we introduce the concept of push-forwards which shift measures from one space onto another via a measurable map.

Definition 2.1.5. *[18, §3.6] Let $(X,\mathcal{A})$ and $(Y,\mathcal{B})$ be two measure spaces and let $T : (X,\mathcal{A}) \to (Y,\mathcal{B})$ be a measurable map. For a measure $\mu \in \mathcal{M}^+(X)$ we define the* ***push - forward of*** *μ* ***under*** *T by*

$$T_{\#}\mu(B) = \mu(T^{-1}(B)) \qquad \forall B \in \mathcal{B}.$$

In particular, $T_{\#}\mu \in \mathcal{M}^+(Y)$.

By a standard approximation of measurable function, one can prove the following change-of-variables formula (cf. Theorem 3.6.1 in [18]):

Proposition 2.1.6. *Under notation of Definition 2.1.5, a measurable map $f : Y \to \mathbb{R}$ is integrable with respect to $T_{\#}\mu$ if and only if the composition $f \circ T$ is integrable with respect to μ. In this the case the following* ***change of variable formula*** *holds*

$$\int_Y f \,\mathrm{d}(T_{\#}\mu) = \int_X f \circ T \,\mathrm{d}\mu. \tag{2.1.1}$$

By choosing f in (2.1.1) to be a characteristic function we see that the push-forward measure $T_{\#}\mu$ is uniquely defined on $(Y,\mathcal{B})$.

Remark 2.1.7. Among other areas of applications, push-forward measures appear in optimal transport theory to represent deterministic couplings of measures under an optimal transport map T, see (see [159, p.16-17]).

Now, we introduce the concept of separating families for subsets of measures. In short, separating families indicate which subclass of test functions already contains enough information to uniquely distinguish all measures of a given set.

Definition 2.1.8. *[94, 13.9] Let $\mathcal{N} \subseteq \mathcal{M}(S)$. A family $\mathcal{F}$ of measurable maps is called* ***separating*** *for $\mathcal{N}$ if for all $\mu, \nu \in \mathcal{N}$ the following implication holds:*

$$\left(\int_S f \,\mathrm{d}\mu = \int_S f \,\mathrm{d}\nu \qquad \forall f \in \mathcal{F} \cap L^1(\mu) \cap L^1(\nu) \right) \Longrightarrow \mu = \nu.$$

In case of the Euclidean space the next result provides a powerful separating family for the space of finite Borel measures.

Proposition 2.1.9. *$C_c^\infty(\mathbb{R}^d)$ is separating for $\mathcal{M}(\mathbb{R}^d)$.*

Proof. Let $\mu, \nu \in \mathcal{M}(\mathbb{R}^d)$ with $\int_{\mathbb{R}^d} f \,\mathrm{d}\mu = \int_{\mathbb{R}^d} f \,\mathrm{d}\nu$ for all $f \in C_c^\infty(\mathbb{R}^d)$. The inner regularity (see e.g. [94, 13.3]) of μ implies that for all $A \in \mathcal{B}(\mathbb{R}^d)$

$$\mu(A) = \sup \{\mu(K) \mid K \subseteq A\}$$

and similarly for ν. In particular, it is sufficient to check equality of μ and ν on compact sets $K \subseteq \mathbb{R}^d$. Let $K \subseteq \mathbb{R}^d$ be compact. According to the C^∞ Urysohn Lemma (see [63, 8.18]) there exists a sequence of functions $(f_n)_{n\in\mathbb{N}} \subset C_c^\infty(\mathbb{R}^d)$ with $0 \leq f_n \leq 1$, $f \equiv 1$ on K, $\operatorname{supp}(f_n) \subseteq B_{1/n}(K)$ and $f_n \to \chi_K$ pointwise. Here

$$B_{1/n}(K) = \left\{ x \in \mathbb{R}^d \mid d(x, K) \leq \frac{1}{n} \right\}$$

denotes the open $1/n$ environment of K. As $\operatorname{supp}(f_n) \subseteq \overline{B_1(K)}$ for all $n \in \mathbb{N}$, we have $f_n \leq \chi_{\overline{B_1(K)}}$ and we can apply Dominated Convergence Theorem, which together with the assumption yields

$$\begin{aligned} \mu(K) &= \int_{\mathbb{R}^d} \chi_K \,\mathrm{d}\mu = \int_{\mathbb{R}^d} \lim_{n\to\infty} f_n \,\mathrm{d}\mu = \lim_{n\to\infty} \int_{\mathbb{R}^d} f_n \,\mathrm{d}\mu \\ &= \lim_{n\to\infty} \int_{\mathbb{R}^d} f_n \,\mathrm{d}\nu = \int_{\mathbb{R}^d} \lim_{n\to\infty} f_n \,\mathrm{d}\nu = \int_{\mathbb{R}^d} \chi_K \,\mathrm{d}\nu = \nu(K). \end{aligned}$$

We conclude $\mu = \nu$ as desired. □

Lemma 2.1.10. *Let $f \in C_c^\infty(\mathbb{R}^d)$ and define $f^\pm := \max\{\pm f, 0\}$. Then there exist sequences of nonnegative functions $(f_n^{(\pm)})_{n\in\mathbb{N}} \subset C_c^\infty(\mathbb{R}^d)$ such that $f_n^{(\pm)} \to f^\pm$*

uniformly. Furthermore, the sequences $(f_n^{(\pm)})_{n\in\mathbb{N}}$ *are uniformly compactly supported and uniformly bounded.*

Proof. By assumption $f \in C_c^\infty(\mathbb{R}^d)$, so f^+ is continuous and compactly supported, i.e. $f^+ \in C_c \subset C_0(\mathbb{R}^d)$. The space $C_c^\infty(\mathbb{R}^d)$ is dense in $\big(C_0(\mathbb{R}^d), \|\cdot\|_\infty\big)$ by definition and thus there exist a sequence $(f_n^{(+)})_{n\in\mathbb{N}} \subset C_c^\infty(\mathbb{R}^d)$ converging uniformly to f^+. As $f^+ \geq 0$ the functions $f_n^{(+)}$ can also be chosen to be nonnegative as the approximating sequence is constructed by convolution with a nonnegative kernel. Due to the uniform convergence we can assume that the sequence is uniformly bounded by

$$\|f_n^{(+)}\|_\infty \leq 2\|f^+\|_\infty \leq 2\|f\|_\infty \qquad \forall n \in \mathbb{N}.$$

Furthermore, the uniform convergence together with the compact support of f^+ implies that we can find a compact set K which contains the supports of all $f_n^{(+)}$, i.e.

$$\mathrm{supp} f_n^{(+)} \subset K \qquad \forall n \in \mathbb{N}.$$

This finishes the proof as the case of f^- is completely similar. $\square$

2.2 The flat norm

On $\mathcal{M}(S)$ there are several norms, the most famous being the total variation norm $\|\cdot\|_{TV}$

$$\|\mu\|_{TV} := |\mu|(S) = \mu^+(S) + \mu^-(S),$$

where $|\cdot|$ denotes the variation of μ and μ^+, μ^- arise from the Hahn-Jordan decomposition. It is well-known that $(\mathcal{M}(S), \|\cdot\|_{TV})$ is a Banach space [44, Theorem 1.20]. If the metric space (S, d) is locally compact and separable, then the total variation norm also has a variational representation [7, Proposition 1.47], i.e. for $\mu \in \mathcal{M}(S)$ and $A \subseteq S$ open

$$|\mu|(A) = \sup\left\{ \int_S \psi \,\mathrm{d}\mu \,\middle|\, \psi \in C_c^0(A), \|\psi\|_\infty \leq 1 \right\}. \tag{2.2.1}$$

Yet, the corresponding topology is too strong and completely ignores the underlying geometry, so that the $\|\cdot\|_{TV}$ norm is not viable for applications. For instance the distance between two Dirac measures is given by

$$\|\delta_a - \delta_b\|_{TV} = \delta_a(S) + \delta_b(S) = 2 \qquad \forall a, b \in S, a \neq b, \tag{2.2.2}$$

so that $\delta_a \nrightarrow \delta_b$ even if $a \to b$. Furthermore, (2.2.2) implies that $\mathcal{M}(S)$ is not separable for uncountable S.

Instead, we will use the weaker **flat norm** (or **bounded Lipschitz norm**, **Fortet-Mourier norm**)

$$\|\mu - \nu\|_{BL^*} := \sup \left\{ \int_S \psi \, \mathrm{d}(\mu - \nu) \mid \psi \in BL(S), \|\psi\|_{BL} \leq 1 \right\} \tag{2.2.3}$$

and the corresponding **(flat) metric** ρ_F. The space of test functions is given by the bounded Lipschitz functions

$$BL(S) = \{f \in C^0(S) \mid \|f\|_{BL} < \infty\},$$

with the (semi-)norms

$$\|f\|_{BL} := \max\left\{\|f\|_\infty, |f|_{\mathbf{Lip}}\right\}, \quad \|f\|_\infty = \sup_{x \in S} |f(x)|, \quad |f|_{\mathbf{Lip}} = \sup_{x \neq y} \frac{|f(x) - f(y)|}{d(x,y)}.$$

Unless stated otherwise, the spaces $\mathcal{M}(S)$ and $\mathcal{M}^+(S)$ will always be equipped with the flat metric and all topological properties refer to the corresponding topology.

In contrast to the $\|\cdot\|_{TV}$ norm, the flat metric respects the geometry of the underlying space, at least locally. This can be seen in the following lemma.

Lemma 2.2.1. *[44, 1.32] Let (S, d) be a metric space. Let $x, y \in S$. Then $\|\delta_x\|_{BL^*} = 1$ and*

$$\rho_F(\delta_x, \delta_y) = \min\{2, d(x,y)\}.$$

In particular, $\delta_x \to \delta_y$ in ρ_F if $x \to y$.

2.3 The space of measures under the flat norm

The space of measures equipped with the flat norm has several convenient properties which will be essential for our generalised model setting in Section 4. We present these aspects in an (almost) self-contained way and highlight the results which will be important later.

Theorem 2.3.1. *Let (S, d) be separable. Then the spaces $\mathcal{M}(S)$ and $\mathcal{M}^+(S)$ are separable with countable dense set given by the linear span* $\operatorname{lin}\{\delta_s\}_{s\in S}$.

We only sketch the main idea of the proof and refer to [44, 1.37] for a complete proof. First, one shows that we can approximate all measures in $\mathcal{M}(S)$ in the flat metric by the linear span of Dirac measures concentrated in the dense subset of S. Then, we restrict the weights to be rational so that the approximation set becomes countable while still being dense. A further restriction to nonnegative weights yields the claim for $\mathcal{M}^+(S)$.
Next, we characterise the predual space of the bounded Lipschitz function which will be helpful for the treatment of the Bochner integrals appearing in Section 4.

Theorem 2.3.2. *Let (S, d) be separable and denote $E := \overline{\mathcal{M}(S)}^{\|\cdot\|_{BL^*}}$. Then it holds that $E^* = BL(S)$.*

Proof. Let $D \subseteq \mathcal{M}(S)$ be the linear space spanned by Dirac deltas, i.e.

$$D := \operatorname{lin}\{\delta_s \mid s \in S\} = \left\{ \sum_{k=1}^{n} \alpha_k \delta_{s_k} \mid n \in \mathbb{N}, \alpha_k \in \mathbb{R}, s_k \in S \right\}.$$

The corresponding (functional analyic) dual space is denoted by D^*. In part one of the proof, we show that we can identify D^* with $BL(S)$. The conclusion will then follow by density of D in $\mathcal{M}(S)$. To prove the claim, for $f \in BL(S)$ define the linear and bounded functional $T_f \in D^*$ via

$$T_f : D \to \mathbb{R}, \qquad T_f(\mu) = \int_S f \, \mathrm{d}\mu.$$

Note that by linearity any functional $T \in D^*$ is uniquely defined by the values $T(\delta_s)$ for $s \in S$ and thus we identify T with the function

$$f : S \to \mathbb{R}, \qquad s \mapsto T(\delta_s).$$

We claim that $f \in BL(S)$. Indeed, if f is not bounded, there exists a sequence $(s_n)_{n\in\mathbb{N}}$ with $\lim\limits_{n\to\infty} f(s_n) = \infty$. But then

$$\lim_{n\to\infty} T(\delta_{s_n}) = \lim_{n\to\infty} f(s_n) = \infty,$$

so the functional can not be bounded as well. On the other hand, if f is not Lipschitz continuous we can assume that there are sequences $(s_n)_{n\in\mathbb{N}}, (t_n)_{n\in\mathbb{N}}$ such that

$$\lim_{n\to\infty} \frac{f(s_n) - f(t_n)}{d(s_n, t_n)} = \infty.$$

Define the sequence of measures $(\mu_n)_{n\in\mathbb{N}}$ with $\mu_n = \frac{\delta_{s_n} - \delta_{t_n}}{d(s_n,t_n)}$ and note that

$$T(\mu_n) = \frac{T(\delta_{s_n}) - T(\delta_{t_n})}{d(s_n, t_n)} = \frac{f(s_n) - f(t_n)}{d(s_n, t_n)} \to \infty.$$

However, for all $n \in \mathbb{N}$

$$\|\mu_n\|_{BL^*} = \sup_{\|\psi\|_{BL}\leq 1} \frac{\psi(s_n) - \psi(t_n)}{d(s_n, t_n)} \leq 1,$$

so T can not be bounded. We conclude that we can indeed identify D^* with $BL(S)$. By Theorem 2.3.1, D is dense in E and we can thus extend any $T \in D^*$ uniquely to a functional in $\tilde{T} \in E^*$ (independent of the approximating sequence) completing the proof. $\square$

Now, we investigate under which conditions the measure spaces are complete with respect to the flat norm as this will be crucial for the application of Banach Fixed Point Theorem. Unfortunately, $\mathcal{M}(S)$ is in general not complete since according to Theorem 1.36 in Ref. [44] the space is only complete iff S is uniformly discrete, i.e. if $\inf_{\substack{x\neq y \\ x,y\in S}} d(x, y) > 0$.

However, we don't want to restrict ourselves to uniformly discrete metric spaces, so we have to find another way for achieving completeness. In the following, we will thus focus on the cone $\mathcal{M}^+(S)$. We start with a theorem which shows that on $\mathcal{M}^+(S)$ narrow convergence is compatible with the topology induced by the flat metric. We omit the proof here, as it is based on even more auxiliary results which would distract from the main part of this section. Instead, we refer to Theorem 1.57 in Ref. [44].

Theorem 2.3.3. *Let (S,d) be a Polish metric space. Then $\|\cdot\|_{BL^*}$ metricizes the narrow topology on $\mathcal{M}^+(S)$, i.e. we have*

$$\mu_n \to \mu \text{ narrowly } \iff \|\mu_n - \mu\|_{BL^*} \to 0.$$

With the help of Theorem 2.3.3 we can formulate a variant of the Prokhorov Theorem for the flat metric.

Theorem 2.3.4. *Let (S,d) be a Polish metric space and let $(\mu_n)_{n\in\mathbb{N}}$ be a tight sequence in $\mathcal{M}^+(S)$ with $\sup\limits_{n\in\mathbb{N}} \mu_n(S) < \infty$. Then $(\mu_n)_{n\in\mathbb{N}}$ has a converging subsequence in $\mu \in \mathcal{M}^+(S)$.*

Proof. Applying the Theorem of Prokhorov 2.1.3 yields a narrowly converging subsequence with limit in $\mathcal{M}^+(S)$. By Theorem 2.3.3 narrow convergence and convergence with respect to the flat metric are equivalent. $\square$

Remark 2.3.5. If S is compact, then Theorem 2.3.4 implies that every bounded and closed subset of $\mathcal{M}^+(S)$ is compact. This property is called **proper**. Proper metric spaces are separable and complete by Proposition A.15 in Ref. [44].

By definition of the flat norm, we have for all $\mu, \nu \in \mathcal{M}^+(S)$ an estimate of the form

$$(\mu-\nu)(S) = \int_S 1 \,\mathrm{d}(\mu-\nu)(x) \leq \|\mu-\nu\|_{BL^*}.$$

Unfortunately, a comparable estimate for arbitrary Borel subsets of S it is in general wrong. Let $\mu = \delta_1$ and $\nu = \delta_{1/2}$. Then according to Lemma 2.2.1

$$\mu(\mathbb{R}) - \nu(\mathbb{R}) = 0 \leq \frac{1}{2} = \|\mu-\nu\|_{BL^*},$$

but for $A = \left(\frac{1}{2}, \infty\right) \in \mathcal{B}(\mathbb{R})$

$$\mu(A) - \nu(A) = 1 - 0 \geq \frac{1}{2} = \|\mu-\nu\|_{BL^*}.$$

However, the following lemma provides pseudo triangle inequality which essentially controls the deviation from the upper bound $\|\mu-\nu\|_{BL*}$.

Lemma 2.3.6. *Let $\delta \in (0,1]$, $T \in \mathcal{B}(S)$. Then for $\mu, \nu \in \mathcal{M}^+(S)$*

$$\mu(T) \leq \nu(B_\delta(T)) + \frac{1}{\delta} \|\mu - \nu\|_{BL^*},$$
$$\mu(S \setminus B_\delta(T)) \leq \nu(S \setminus T) + \frac{1}{\delta} \|\mu - \nu\|_{BL^*}.$$

Here $B_\delta(T) := \{x \in S \mid d(x,T) < \delta\}$ denotes the δ-neighborhood of T.

Proof. Consider the function $\eta : S \to [0,1]$

$$\eta(x) = \frac{d(x, S \setminus B_\delta(T))}{d(x, S \setminus B_\delta(T)) + d(x,T)}.$$

According to Lemma 2.1 in Ref. [81], η is Lipschitz continous with Lipschitz constant bounded by $1/\delta$, so that $\|\eta\|_{BL} \leq 1/\delta$. Furthermore, by construction $\eta \equiv 1$ on T and $\eta \equiv 0$ on $S \setminus B_\delta(T)$. Thus,

$$\mu(T) \leq \int_S \eta \,\mathrm{d}\mu \leq \frac{1}{\delta} \|\mu - \nu\|_{BL^*} + \int_S \eta \,\mathrm{d}\nu \leq \frac{1}{\delta} \|\mu - \nu\|_{BL^*} + \nu(B_\delta(T)).$$

For the other estimate, we switch the roles of T and $S \setminus B_\delta(T)$. $\square$

Proposition 2.3.7. *Let (S,d) be a Polish metric space. If $\mathcal{N} \subseteq \mathcal{M}^+(S)$ is totally bounded with respect to $\|\cdot\|_{BL^*}$, then $\mathcal{N}$ is tight.*

Proof. The main idea of this proof is based on the reasoning from Ref. [81] but with more explanations to increase the readability. Let $\mathcal{N} \subseteq \mathcal{M}^+(S)$ be a totally bounded set of measures. In a first step, we discretize $\mathcal{N}$ using the total boundedness, i.e. for $\varepsilon \in (0,1)$, there exists a finite subset $\tilde{\mathcal{N}} \subseteq \mathcal{N}$ such that

$$\mathcal{N} \subseteq \bigcup_{\nu \in \tilde{\mathcal{N}}} B_{\varepsilon^2/4}(\nu) =: B_{\varepsilon^2/4}(\tilde{\mathcal{N}}). \tag{2.3.1}$$

As S is a Polish metric space, $\{\mu\}$ is tight for any finite measure μ, cf. Remark 13.27 in Ref. [94], so that the finite set $\tilde{\mathcal{N}}$ is tight. In particular, there exists a compact set $K \subseteq S$ such that $\nu(S \setminus K) < \varepsilon/2$ for all $\nu \in \tilde{\mathcal{N}}$. Next, we discretize K. As K is compact, we cover K with finitely many sets of the form $B_{\varepsilon/2}(x_i)$ with $x_i \in F = F_\varepsilon \subset K$. Note that still

$$\nu(S \setminus B_{\varepsilon/2}(F)) \leq \nu(S \setminus K) < \varepsilon/2 \qquad \forall \nu \in \tilde{\mathcal{N}}.$$

Let $\mu \in \mathcal{N}$ be arbitrary with corresponding approximating measure $\nu \in \tilde{\mathcal{N}}$ such that $\|\mu - \nu\|_{BL^*} \leq \varepsilon^2/4$. Triangle inequality implies $B_{\varepsilon/2}(B_{\varepsilon/2}(F)) \subseteq B_\varepsilon(F)$, so that with Lemma 2.3.6

$$\mu(S \setminus B_\varepsilon(F)) \leq \mu(S \setminus B_{\varepsilon/2}(B_{\varepsilon/2}(F))) \leq \nu(S \setminus B_{\varepsilon/2}(F)) + \frac{2}{\varepsilon} \|\mu - \nu\|_{BL^*} < \varepsilon.$$

In particular, for each $\varepsilon \in (0,1)$ we constructed a finite set $F_\varepsilon \subseteq S$ such that $\mu(S \setminus B_\varepsilon(F_\varepsilon)) < \varepsilon$ for all $\mu \in \mathcal{N}$.

We conclude the proof by using the family of sets $\{F_\varepsilon\}_{\varepsilon<1}$ to construct a suitable compact set to prove tightness of $\mathcal{N}$. Let $\delta \in (0,1]$ be arbitrary and for $n \in \mathbb{N}$ set $\varepsilon_n = \delta 2^{-n}$. With the considerations above we obtain for each $n \in \mathbb{N}$ a finite set $F_n := F_{\varepsilon_n}$ such that

$$\mu(S \setminus B_{\varepsilon_n}(F_n)) < \varepsilon_n \quad \forall \mu \in \mathcal{N}.$$

The infinite intersection of closed sets $K_\delta := \bigcap_{n\in\mathbb{N}} \overline{B_{\varepsilon_n}(F_n)}$ is closed and totally bounded by construction, i.e. compact. Furthermore, we have for all $\mu \in \mathcal{N}$

$$\mu(S \setminus K_\delta) \leq \mu\left(\bigcup_{n\in\mathbb{N}} S \setminus B_{\varepsilon_n}(F_n)\right) \leq \sum_{n=1}^{\infty} \mu(S \setminus B_{\varepsilon_n}(F_n)) < \sum_{n=1}^{\infty} \varepsilon_n = \delta \sum_{n=1}^{\infty} 2^{-n} = \delta.$$

In particular, $\mathcal{N}$ is tight. □

Combining the above results, we can prove the desired completeness of the cone.

Theorem 2.3.8. *Let (S, d) be a Polish metric space. Then $\mathcal{M}^+(S)$ is complete with respect to the flat norm.*

Proof. Let $(\mu_n)_{n\in\mathbb{N}}$ be a Cauchy sequence. Then the set $\mathcal{N} := \{\mu_n \mid n \in \mathbb{N}\}$ is totally bounded and we conclude tightness of $\mathcal{N}$ from Proposition 2.3.7. Since $(\mu_n)_{n\in\mathbb{N}}$ is Cauchy, the real valued sequence $(\mu_n(S))_{n\in\mathbb{N}}$ is bounded, so that we can apply Theorem 2.3.4 to extract a converging subsequence. However, a Cauchy sequence with a converging subsequence is already convergent which was to show. □

We conclude this subsection with a different characterisation of the cone $\mathcal{M}^+(S)$ by the means of positive functionals on $BL(S)$. These functionals appear naturally in limiting procedures such as the construction of the Bochner integral on $E = \overline{\mathcal{M}(S)}^{\|\cdot\|_{BL^*}}$ in Proposition 4.2.4. In general, $\mathcal{M}(S) \subsetneq E$ so not all elements in E are measures but only abstract elements in the closure. Fortunately, according to a result from [87] it turns out that positive functionals are actually nonnegative measures.

Theorem 2.3.9. *Consider the space of* ***positive linear functionals*** *on* $BL(S)$ *given by* $BL(S)^*_+ := \{T \in BL(S)^* \mid T(\psi) \geq 0 \, \forall \psi \in BL(S), \psi \geq 0\}$*. Then it holds that*

$$E \cap BL(S)^*_+ = \mathcal{M}^+(S). \tag{2.3.2}$$

Proof. We follow the reasoning from Ref. [87] but streamline the ideas. First, note that any $T \in BL^*_+(S)$ is monotone, i.e.

$$T(\psi) \geq T(\varphi) \qquad \forall\, \psi, \varphi \in BL(S) \text{ with } \psi \geq \varphi,$$

which follows directly from the positivity and the linearity of the operator.
We show (2.3.2). It is directly clear that $\mathcal{M}^+(S) \subseteq E \cap BL(S)^*_+$. For the other inclusion, assume by way of contradiction that there is $T \in E \cap BL(S)^*_+$ such that $T \notin \mathcal{M}^+(S)$. Let $\mathbf{1} \in BL(S)$ be the constant function with value 1. We claim that if T vanishes for $\mathbf{1}$, then it must already vanish for all $\psi \in BL(S)$. Indeed, as T is linear and monotone, we compute

$$\begin{aligned} |T(\psi)| &= |T(\psi^+) - T(\psi^-)| \leq |T(\psi^+)| + |T(\psi^-)| = T(\psi^+) + T(\psi^-) \\ &= T(\psi^+ + \psi^-) = T(|\psi|) \leq T(\|\psi\|_\infty \cdot \mathbf{1}) = \|\psi\|_\infty T(\mathbf{1}) = 0, \end{aligned}$$

where $\psi^\pm = \max\{\pm\psi, 0\}$. Hence, $T = 0 \in \mathcal{M}^+(S)$, which is a contradiction. Hence, for the remainder of the proof we assume $T(\mathbf{1}) > 0$. Theorem 2.3.8 implies that $\mathcal{M}^+(S)$ is complete with respect to the flat norm, so that in particular $\mathcal{M}^+(S)$ is a closed subset of E. We apply the Hyperplane Separation Theorem to the closed and convex set $\mathcal{M}^+(S)$ and strictly separate it from $\{T\} \nsubseteq \mathcal{M}^+(S)$ in E. Consequently, there is $f \in E^* = BL(S)$ (see Theorem 2.3.2) and $\alpha \in \mathbb{R}$ such that

$$\langle \mu, f \rangle \leq \alpha \quad \forall \mu \in \mathcal{M}^+(S) \quad \text{and} \quad \langle T, f \rangle = T(f) > \alpha. \tag{2.3.3}$$

As $T(\mathbf{1})\delta_x \in \mathcal{M}^+(S)$ for all $x \in S$, we get

$$\alpha \geq \langle T(\mathbf{1})\delta_x, f \rangle = T(\mathbf{1}) f(x) \text{ for all } x \in S$$

and thus, $f \leq \frac{\alpha}{T(\mathbf{1})}$. Using the positivity of T

$$T(f) \leq T\left(\frac{\alpha\mathbf{1}}{T(\mathbf{1})}\right) = \alpha$$

yields a contradiction to (2.3.3) and hence $E \cap BL(S)^*_+ = \mathcal{M}^+(S)$. □

2.4 The flat norm as a distance for unbalanced optimal transport

In this section, we will explain that the flat norm has some remarkable similarities with the Wasserstein metrics known from *optimal transport* theory. In particular, the definition of the flat metric (2.2.3) resembles the Kantorovich-Rubinstein duality of the Wasserstein distance W_1, i.e.

$$W_1(\mu, \nu) = \sup_{|f|_{\mathbf{Lip}} \leq 1} \int_S f \, \mathrm{d}(\mu - \nu). \tag{2.4.1}$$

The Wasserstein metrics define distances between probability measures which take into account the geometry of the underlying state space [33, 158, 159]. Consequently, distances with respect to the Wasserstein metrics are more robust which makes it a suitable candidate for various machine learning tasks [76, 77, 102, 118]. The Wasserstein distances scale with the total mass of the measures μ, ν and are thus not necessarily restricted to probability measures. However, by construction the distances are only applicable in conservative problems, i.e. only if $\mu(S) = \nu(S)$, as otherwise no optimal transport plan exists, see e.g. [156, Remark 1.18].
In cases where the measures or data distributions can not be normalised as the mass differences are meaningful, see e.g. color transfer, other approaches are necessary to handle these so-called *unbalanced optimal transport* tasks. See [28, 117] for an overview of several approaches on unbalanced optimal transport. As the flat metric ρ_F defines a distance on $\mathcal{M}^+(S)$, it is a suitable candidate for this kind of problem as well. This claim is supported by the following result, which provides an alternative characterisation of the flat metric for spaces with more structure via a *primal formulation*- in contrast to the *dual formulation* (2.2.3). The original result of Piccoli and Rossi in the case $S = \mathbb{R}^d$ (cf. [120, Theorem 13]) can be extended to proper metric spaces.

Theorem 2.4.1. *Let (S,d) be a proper metric space and $\mu,\nu \in \mathcal{M}_1^+(S)$. Then*

$$\rho_F(\mu,\nu) = \inf_{\substack{\tilde{\mu}\le\mu,\,\tilde{\nu}\le\nu \\ \|\tilde{\mu}\|_{TV}=\|\tilde{\nu}\|_{TV}}} \|\mu-\tilde{\mu}\|_{TV} + \|\nu-\tilde{\nu}\|_{TV} + W_1(\tilde{\mu},\tilde{\nu}). \tag{2.4.2}$$

Here, W_1 denotes the classical $1-$Wasserstein distance with respect to the cost function $c(x,y) = d(x,y)$.

Remark 2.4.2. The decomposition into terms with total variation norm and the term with Wasserstein distance admits a heuristic interpretation of mass deletion versus mass transportation: Any share $\delta\mu$ of the mass of μ can either be transported from μ to ν at cost $W_1(\delta\mu,\delta\nu)$ or removed at cost $\|\delta\mu\|_{TV}$. As such, the minimal "sub-measures" $\tilde{\mu},\tilde{\nu}$ achieve an optimal compromise between mass transportation and cancellation.

Remark 2.4.3. Theorem 2.4.1 shows that in a way the flat metric is a suitable generalisation of the 1-Wasserstein distance W_1 for unbalanced tasks and as such distances with respect to ρ_F are also geometrically faithful, at least locally (see Lemma 2.2.1).

We adapt the proof from [121] for the $\mathbb{R}^d$ case to proper metric spaces. To this end, we first recall some definitions and results on convex functionals which will appear in the proof.

Definition 2.4.4. *Let X be a Banach space and $F : X \to \overline{\mathbb{R}}$ a function. The* ***conjugate function*** *$F^* : X^* \to \overline{\mathbb{R}}$ is defined by*

$$F^*(y) := \sup_{x\in X} \left(\langle y,x\rangle - F(x)\right),$$

where $\langle\cdot,\cdot\rangle$ denotes the usual dual pairing.

Definition 2.4.5. *Let X be a Banach space. A function $F : X \to \overline{\mathbb{R}}$ is called* ***closed*** *if the sublevel sets $\{x \in X \mid F(x) \le k\}$ are closed in X for all $k \in \overline{\mathbb{R}}$.*

The following technical theorem allows us to efficiently compute the conjugate function of a sum of two functions. For more details and a proof see [129, Theorem 20.e].

Theorem 2.4.6. *Let X be a Banach space and let $F_1, F_2 : X \to \mathbb{R}\cup\{\infty\}$ be convex and closed. Assume that there exists a neighbourhood U of $0 \in X$, an open set*

$M \subseteq X^*$ and a constant K such that for all sets

$$V_\alpha := \{(y_1, y_2) \mid y_i \in \operatorname{dom} F_i^*, y_1 + y_2 \in M, F_1^*(y_1) + F_2^*(y_2) \leq \alpha\}$$

it holds that

$$\sup_{x \in U,\, (y_1, y_2) \in V_\alpha} \langle y_1 + y_2, x \rangle < K.$$

Here $\operatorname{dom} F_i^*$ *denotes the domain of F_i^*. Then the conjugate function F^* of $F = F_1 + F_2$ satisfies*

$$F^*(y) = \min_{y_1 + y_2 = y} (F_1^*(y_1) + F_2^*(y_2)).$$

Proof of Theorem 2.4.1. Consider the following functionals on $X = (C_0(S), \|\cdot\|_\infty)$

$$F_1(\psi) := \begin{cases} 0, & \|\psi\|_\infty \leq 1 \\ +\infty, & \text{else} \end{cases} \qquad F_2(\psi) := \begin{cases} 0, & |\psi|_{\mathbf{Lip}} \leq 1 \\ +\infty, & \text{else} \end{cases}.$$

As proper spaces are locally compact and separable, the Riesz Representation Theorem [62, 7.17] implies that

$$\mathcal{M}(S) = C_0(X)^* = X^*.$$

Note that by the dual formulation of the total variation norm (2.2.1) and as $C_c^0(S)$ is dense in $C_0(S)$ we have

$$\begin{aligned} F_1^*(\mu - \nu) &= \sup_{\psi \in X} \left(\int_S \psi \, \mathrm{d}(\mu - \nu) - F_1(\psi) \right) = \sup_{\psi \in X} \left\{ \int_S \psi \, \mathrm{d}(\mu - \nu) \,\middle|\, \|\psi\|_\infty \leq 1 \right\} \\ &= \sup \left\{ \int_S \psi \, \mathrm{d}(\mu - \nu) \,\middle|\, \|\psi\|_\infty \leq 1, \psi \in C_0(S) \right\} = \|\mu - \nu\|_{TV}. \end{aligned} \tag{2.4.3}$$

For the Wasserstein distance there exists an analogous dual formulation which is a variant of the Kantorovich-Rubinstein duality for proper metric spaces (see [44, 1.85]), i.e. for $\mu, \nu \in \mathcal{M}_1^+(S)$ with $\|\mu\|_{TV} = \|\nu\|_{TV}$ it holds that

$$W_1(\mu, \nu) = \sup \left\{ \int_S \psi \, \mathrm{d}(\mu - \nu) \,\middle|\, \psi \in C_c^0, |\psi|_{\mathbf{Lip}} \leq 1 \right\}. \tag{2.4.4}$$

Similarly to (2.4.3), using (2.4.4) we see that $F_2^*(\mu-\nu) = W_1(\mu,\nu)$.

In the next step we define $F := F_1 + F_2$ and see by the definition of the conjugate function that

$$F^*(\mu-\nu) = \sup_{\psi\in X}\left\{\int_S \psi\,\mathrm{d}(\mu-\nu)\,\middle|\, \|\psi\|_\infty \le 1, |\psi|_{\mathbf{Lip}} \le 1\right\} = \rho_F(\mu,\nu).$$

Note that we used [44, 1.26] in the last step which states that for S proper the space of test functions for the flat metric can be restricted to $BL(S)\cap C_c^0(S)$.
We conclude the proof by proving that $F^*(\mu-\nu)$ is actually the RHS of (2.4.2).
By construction, the functions F_i are closed and convex. Furthermore, $F_i(\psi) > -\infty$ for all $\psi\in X$ and $F_i(X)\neq\emptyset$, i.e. the F_i are proper in the sense of [129]. We choose

$$U := \{\psi\in X \mid \|\psi\|_\infty < \varepsilon\}, \qquad M = \{\mu\in\mathcal{M}(S) \mid \|\mu\|_{TV} < \varepsilon\}$$

and $K = \varepsilon^2$. Then the conditions of Theorem 2.4.6 are satisfied, i.e. in the notation of the theorem we have for

$$(\mu_1,\mu_2)\in V_\alpha = \{(\nu_1,\nu_2) \mid \|\nu_1+\nu_2\|_{TV} < \varepsilon, F_1^*(\nu_1) + F_2^*(\nu_2) \le \alpha\}$$

that

$$\sup_{\substack{\psi\in U,\\ (\mu_1,\mu_2)\in V_\alpha}} \langle \mu_1+\mu_2,\psi\rangle = \sup_{\substack{\|\psi\|_\infty<\varepsilon,\\ (\mu_1,\mu_2)\in V_\alpha}} \int_{\mathbb{R}^n}\psi\,\mathrm{d}(\mu_1+\mu_2) \le \|\psi\|_\infty\|\mu_1+\mu_2\|_{TV} \le \varepsilon^2.$$

Applying Theorem 2.4.6 thus concludes the proof

$$\begin{aligned}
F^*(\mu-\nu) &= \min_{\substack{(\mu_1-\nu_1)+(\mu_2-\nu_2)=\mu-\nu\\ \|\mu_2\|_{TV}=\|\nu_2\|_{TV}}} \left(F_1^*(\mu_1-\nu_1) + F_2^*(\mu_2-\nu_2)\right)\\
&= \min_{\substack{(\mu_1-\nu_1)+(\mu_2-\nu_2)=\mu-\nu\\ \|\mu_2\|_{TV}=\|\nu_2\|_{TV}}} \left(\|\mu_1-\nu_1\|_{TV} + W_1(\mu_2,\nu_2)\right)\\
&= \min_{\substack{\tilde\mu\le\mu,\,\tilde\nu\le\nu\\ \|\tilde\mu\|_{TV}=\|\tilde\nu\|_{TV}}} \left(\|\mu-\tilde\mu-(\nu-\tilde\nu)\|_{TV}\right) + W_1(\tilde\mu,\tilde\nu)\\
&= \min_{\substack{\tilde\mu\le\mu,\,\tilde\nu\le\nu\\ \|\tilde\mu\|_{TV}=\|\tilde\nu\|_{TV}}} \left(\|\mu-\tilde\mu\|_{TV} + \|\nu-\tilde\nu\|_{TV}\right) + W_1(\tilde\mu,\tilde\nu).
\end{aligned}$$

Note that we used the additivity of $\|\cdot\|_{TV}$ on $\mathcal{M}^+(S)$ and set $\mu_2 = \tilde\mu, \nu_2 = \tilde\nu$. In particular, as $W_1(\mu_2,\nu_2) < \infty$ we deduced $\|\mu_2\|_{TV} = \|\nu_2\|_{TV}$ and thus $\tilde\mu$ and $\tilde\nu$ have the same mass. □

In Lemma 2.2.1 the flat metric becomes constant if the supports of the Dirac measures are too far apart. In view of Remark 2.4.2 this corresponds to pure mass generation/ removal without any transport. This observation also holds for general measures.

Proposition 2.4.7. *Let $\mu, \nu \in \mathcal{M}^+(S)$ with $\operatorname{dist}(\operatorname{supp}(\mu), \operatorname{supp}(\nu)) \geq 2$. Then*

$$\rho_F(\mu, \nu) = \|\mu\|_{TV} + \|\nu\|_{TV}.$$

Proof. On the one hand, we have for all $\|f\|_{BL} \leq 1$

$$\int_S f \, \mathrm{d}(\mu - \nu) \leq \int_S 1 \, \mathrm{d}|\mu - \nu| = \|\mu - \nu\|_{TV} = \|\mu\|_{TV} + \|\nu\|_{TV},$$

where we used properties of the variation of a measure and of the total variation norm. On the other hand, as $\operatorname{dist}(\operatorname{supp}(\mu), \operatorname{supp}(\nu)) \geq 2$ we can consider a test function f^* which satisfies $f^* \equiv 1$ on $\operatorname{supp}(\mu)$, $f^* \equiv -1$ on $\operatorname{supp}(\nu)$ and $\|f^*\|_{BL} \leq 1$. Such a function can be found for example by linearly interpolating between both supports. Note that it is crucial for the construction of f^* that the supports of both measures are sufficiently distinct. With this choice of f^* we compute

$$\int_S f \, \mathrm{d}(\mu - \nu) = \int_{\operatorname{supp}(\mu)} 1 \, \mathrm{d}\mu - \int_{\operatorname{supp}(\nu)} -1 \, \mathrm{d}\nu = \|\mu\|_{TV} + \|\nu\|_{TV}$$

concluding the proof. □

We conclude this subsection with a generalisation of Lemma 2.2.1 to more elaborate combinations of Dirac deltas.

Proposition 2.4.8. *Let $N \in \mathbb{N}$ and $c \in \mathbb{R}^+$. Consider points $x_0, x_1, ..., x_N \in S$ which are ordered with increasing distance to x_0, i.e. for $d_i := d(x_0, x_i)$ we have $d_1 \leq d_2 \leq \cdots \leq d_N$ and let $l \in \{0, \ldots, N\}$ be such that*

$$\begin{cases} d_i \leq 2, & i \leq l, \\ d_i > 2, & i > l \end{cases}.$$

Furthermore, for $i = 1, ..., n$ consider weights $b_i \in \mathbb{R}^+$. Then the flat distance *between the measures* $\mu = c\delta_{x_0}$ *and* $\nu = \sum_{i=1}^{N} b_i \delta_{x_i}$ *is given by*

$$\rho_F(\mu,\nu) = \sum_{i=1}^{I^*} b_i d_i + \left(\min\left\{c, \sum_{i=1}^{l} b_i\right\} - \sum_{i=1}^{I^*} b_i\right) d_{I^*+1} + \left|c - \sum_{i=1}^{l} b_i\right| + \sum_{i=l+1}^{N} b_i, \tag{2.4.5}$$

where

$$I^* = \max\left\{ I \in \{0, ..., l\} \mid \sum_{i=1}^{I} b_i \leq c \right\}.$$

To clarify formula (2.4.5), we explain the different terms and provide two examples before we start with the proof. From Theorem 2.4.1 we see that the cost of transportation is directly related to the covered distance, so that an optimal-transport-mass-deletion plan should favour closer points for transportation. However, in view of Proposition 2.4.7 mass transport from μ, i.e. from the point x_0, to some point $x_i \in \text{supp}(\nu)$ is only a good strategy, if $d(x_0, x_i) \leq 2$. In any other case, mass deletion/ creation is cheaper and should be preferred. Thus, we are specifically interested in the share of mass of ν which is located within this distance to x_0, marked by the first l support points. However, due to mass imbalances within the circle of radius 2 it can happen that additional mass has to be created or superfluous mass needs to be deleted, even though transport would be more efficient. After all, mass transportation needs both enough mass at the starting point and at the designated end point(s). The auxiliary constant I^* classifies the support points of ν, which together have just less mass than we can send from μ. The next closest point x_{I^*+1} does either not receive any mass via transportation or just a fraction of the desired mass with the rest being assigned by mass generation. All these effects are reflected in formula (2.4.5) as follows:

- The last term $\sum_{i=l+1}^{N} b_i$ denotes the part of the mass of ν which lies outside the efficient distance of transportation and thus has to be created in any case.
- With the term $\left|c - \sum_{i=1}^{l} b_i\right|$ we compute the mass imbalance inside the circle of radius 2. This mass has either to be created as μ does not provide enough mass to start with or is superfluous and thus needs to be deleted.
- The sum $\sum_{i=1}^{I^*} b_i d_i$ represents the mass which will be completely transported from x_0 to points $x_i \in \text{supp}(\nu)$ within the range of efficient transportation. In other words, the support points $x_1, ..., x_{I^*}$ receive their mass exclusively via

transportation. If there is still mass remaining, this is taken into account by the second term in (2.4.5).

The most complicated term is given by $\left(\min\left\{c, \sum_{i=1}^{l} b_i\right\} - \sum_{i=1}^{I^*} b_i\right) d_{I^*+1}$ and clarifies the mass imbalance between μ and ν within the circle of radius 2. If μ outweighs ν inside the circle (i.e. if $c \geq \sum_{i=1}^{l} b_i$), then $I^* = l$ and the whole term vanishes. In this case, the superfluous mass is deleted and the cost have already been accounted for completely by the term $\left|c - \sum_{i=1}^{l} b_i\right|$. However, if ν outweighs μ inside the circle (i.e. $\sum_{i=1}^{l} b_i \geq c$), then we transport the remaining mass of μ, given by

$$c - \sum_{i=1}^{I^*} b_i$$

to the next closest point x_{I^*+1} within the circle. By construction, this amount of mass is smaller than b_{I^*+1}, so that part of the mass at x_{I^*+1} has to be created. We provide two simple example which will further illustrate the application of formula (2.4.5).

Example 2.4.9.

- Let $\mu = 5\delta_0$ and $\nu = \sum_{i=1}^{7} \delta_{x_i}$, where $d_i < 2$ for $i = 1, ..., 4$ and $d_i > 2$ for $i = 5, 6, 7$. See Figure 2.1. Then $c = 5$, $l = 4$, $b_i = 1$ for $i = 1, .., 7$ and $I^* = 4$. Consequently, formula (2.4.5) reads

$$\begin{aligned}\rho_F(\mu, \nu) &= \sum_{i=1}^{4} d_i + \left(\min\left\{5, \sum_{i=1}^{4} 1\right\} - \sum_{i=1}^{4} 1\right) d_5 + \left|5 - \sum_{i=1}^{4} 1\right| + \sum_{i=5}^{7} 1 \\ &= \sum_{i=1}^{4} d_i + 1 + 3 = \sum_{i=1}^{4} d_i + 4.\end{aligned}$$

- Let $\mu = 5/2\delta_0$ and $\nu = \sum_{i=1}^{4} \delta_{x_i}$, where $d_i < 2$ for $i = 1, 2, 3$ and $d_4 > 2$. See Figure 2.2. In this case, $c = 5/2$, $l = 3$, $b_i = 1$ for $i = 1, ..., 4$ and $I^* = 2$. Thus, formula (2.4.5) yields

$$\begin{aligned}\rho_F(\mu, \nu) &= \sum_{i=1}^{2} d_i + \left(\min\left\{\frac{5}{2}, \sum_{i=1}^{3} 1\right\} - \sum_{i=1}^{2} 1\right) d_3 + \left|\frac{5}{2} - \sum_{i=1}^{3} 1\right| + \sum_{i=4}^{4} 1 \\ &= \sum_{i=1}^{2} d_i + \left(\frac{5}{2} - 2\right) d_3 + \frac{1}{2} + 1 = \sum_{i=1}^{2} d_i + \frac{1}{2} d_3 + \frac{3}{2}.\end{aligned}$$

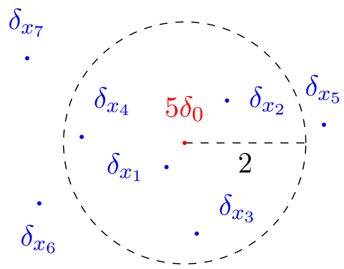

Figure 2.1: $c = 5$, $l = 4$, $N = 7$, $I^* = 4$

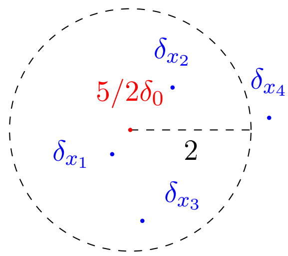

Figure 2.2: $c = 5/2$, $l = 3$, $N = 4$, $I^* = 2$

Proof of Proposition 2.4.8. We apply the alternative characterisation of Piccoli and Rossi from Theorem 2.4.1 and consequently have to find the optimal submeasures $\tilde{\mu}, \tilde{\nu}$. As μ is a Dirac measure located at x_0, any submeasure of μ is of the form $\tilde{\mu}_\alpha = \alpha\delta_{x_0}$ for some $\alpha \in \left[0, \min\left\{c, \sum_{i=1}^N b_i\right\}\right]$. The parameter α denotes the share of the mass located at x_0 which we want to transport to ν, and α is thus à priori bounded by the minimum of the total masses of μ and ν. We directly compute

$$\|\mu - \tilde{\mu}_\alpha\|_{TV} = (c - \alpha)\|\delta_{x_0}\|_{TV} = c - \alpha. \tag{2.4.6}$$

As ν is a linear combination of Diracs, its submeasures $\tilde{\nu}$ are slightly more difficult. However, any $\tilde{\nu}$ is definitely of the form $\tilde{\nu}_{\alpha,\beta} = \sum_{i=1}^N \beta_i\delta_{x_i}$ with weights $\beta_i \in [0, b_i]$ satisfying $\sum_{i=1}^N \beta_i = \alpha$ as both submeasures need to have the same total mass. The weights β_i indicate how much of the mass located at x_i comes via transportation from x_0, whereas the rest $b_i - \beta_i$ has to be created. We compute

$$\|\nu - \tilde{\nu}_{\alpha,\beta}\|_{TV} = \sum_{i=1}^N (b_i - \beta_i)\|\delta_{x_i}\|_{TV} = \sum_{i=1}^N (b_i - \beta_i) = \sum_{i=1}^N b_i - \alpha, \tag{2.4.7}$$

where we used that the total variation norm behaves linearly for nonnegative measures and that $\|\delta_x\|_{TV} = 1$.

Next, we need to check the Wasserstein distance between $\tilde{\mu}_\alpha$ and $\tilde{\nu}_{\alpha,\beta}$. If $\alpha = 0$, then clearly $W_1(\tilde{\mu}_\alpha, \tilde{\nu}_{\alpha,\beta}) = 0$, so let $\alpha > 0$ for the upcoming computation

$$W_1(\tilde{\mu}_\alpha, \tilde{\nu}_{\alpha,\beta}) = W_1\left(\alpha\delta_{x_0}, \sum_{i=1}^N \beta_i\delta_{x_i}\right) = \int_S d(x_0, y)\,\mathrm{d}\left[\sum_{i=1}^N \beta_i\delta_{x_i}\right](y) = \sum_{i=1}^N \beta_i d(x_0, x_i). \tag{2.4.8}$$

Here we used that the Wasserstein distance between a Dirac measure and an arbitrary probability measure η is given by

$$W_1(\eta, \delta_{x_0}) = \int_S d(x_0, y)\, \mathrm{d}\eta(y).$$

In view of identity (2.4.2), we combine (2.4.6), (2.4.7) and (2.4.8) to get the following estimate

$$\begin{aligned} \rho_F(\mu, \nu) \leq& \|\mu - \tilde{\mu}_\alpha\|_{TV} + \|\nu - \tilde{\nu}_{\alpha,\beta}\|_{TV} + W_1(\tilde{\mu}_\alpha, \tilde{\nu}_{\alpha,\beta}) \\ =& c + \sum_{i=1}^{N} b_i - 2\alpha + \sum_{i=1}^{N} \beta_i d_i =: F(\alpha, \beta), \end{aligned} \tag{2.4.9}$$

with $\alpha = \sum_{i=1}^{N} \beta_i$ so that the case $\alpha = 0$ is included. As all possible submeasures are of the form $\tilde{\mu}_\alpha, \tilde{\nu}_{\alpha,\beta}$ we reduced the problem to minimizing F subject to the constraint $\alpha = \sum_{i=1}^{N} \beta_i$. We claim that the global minimum is attained in (α^*, β^*) where

$$\alpha^* = \min\left\{c, \sum_{i=1}^{l} b_i\right\} \qquad \text{and} \qquad \beta_i^* = \begin{cases} b_i, & i \leq I^*, \\ \alpha^* - \sum_{i=1}^{I^*} b_i, & i = I^* + 1, \\ 0, & \text{else} \end{cases} \tag{2.4.10}$$

and $I^* = \max\left\{I \in \{0, ..., l\} \mid \sum_{i=1}^{I} b_i \leq c\right\}$. In Remark 2.4.10 we give a heuristic for this specific parameter choice. Before we prove the optimality, we show that (2.4.10) leads to (2.4.5). So we plug in (α^*, β^*) into F and treat both cases of α^* separately.

<u>Case 1: $\alpha^* = c$</u>

In this case $c \leq \sum_{i=1}^{l} b_i \leq \sum_{i=1}^{N} b_i$, so that

$$\begin{aligned} F(\alpha^*, \beta^*) =& c + \sum_{i=1}^{N} b_i - 2c + \sum_{i=1}^{N} \beta_i^* d_i = \sum_{i=1}^{N} b_i - c + \sum_{i=1}^{I^*} b_i d_i + \left(c - \sum_{i=1}^{I^*} b_i\right) d_{I^*+1} \\ =& \sum_{i=1}^{I^*} b_i d_i + \left(c - \sum_{i=1}^{I^*} b_i\right) d_{I^*+1} + \left|c - \sum_{i=1}^{l} b_i\right| + \sum_{i=l+1}^{N} b_i \end{aligned}$$

as claimed.

<u>Case 2: $\alpha^* = \sum_{i=1}^l b_i$</u>
In this case we have $I^* = l$ so that

$$\begin{aligned} F(\alpha^*, \beta^*) =& c + \sum_{i=1}^N b_i - 2\sum_{i=1}^l b_i + \sum_{i=1}^N \beta_i^* d_i \\ =& c + \sum_{i=1}^N b_i - 2\sum_{i=1}^l b_i + \sum_{i=1}^{I^*} b_i d_i + \left(\alpha^* - \sum_{i=1}^l b_i\right) d_{I^*+1} \\ =& \sum_{i=1}^{I^*} b_i d_i + \left| c - \sum_{i=1}^l b_i \right| + \sum_{i=l+1}^N b_i \end{aligned}$$

as desired.
We are left to prove that (α^*, β^*) yields the global minimum of F. To show this, we invoke variational inequality theory. First note that the domain of F

$$X := \left\{ (\alpha, \beta) \in [0, c] \times \bigotimes_{i=1}^N [0, b_i] \mid \alpha - \sum_{i=1}^N \beta_i = 0 \right\} \subset \mathbb{R}^{N+1}$$

is nonempty, closed and convex. Furthermore, $F : X \to \mathbb{R}$ is smooth, linear and thus convex. Let $x^* := (\alpha^*, \beta^*)$. According to [70, 7.5] the point x^* is a global minimum of F if x^* solves the variational inequality

$$VIP(X, \nabla F) := \nabla F(x^*)^T (x - x^*) \geq 0 \qquad \forall x \in X. \tag{2.4.11}$$

We compute for some $x = (\alpha, \beta) \in X$

$$\nabla F(x^*)^T (x - x^*) = \begin{pmatrix} -2 \\ d_1 \\ \vdots \\ d_N \end{pmatrix}^T \begin{pmatrix} \alpha - \alpha^* \\ \beta_1 - \beta_1^* \\ \vdots \\ \beta_N - \beta_N^* \end{pmatrix} = -2(\alpha - \alpha^*) + \sum_{i=1}^N d_i(\beta_i - \beta_i^*)$$

$$= -2\left(\sum_{i=1}^N \beta_i - \sum_{i=1}^N \beta_i^*\right) + \sum_{i=1}^N d_i(\beta_i - \beta_i^*) = \sum_{i=1}^N (d_i - 2)(\beta_i - \beta_i^*).$$

Plugging in β^* yields

$$\begin{aligned}&\nabla F(x^*)^T(x-x^*)\\&=\sum_{i=1}^{I^*}(d_i-2)(\beta_i-b_i)+(d_{I^*+1}-2)\left(\beta_{I^*+1}-\alpha^*+\sum_{i=1}^{I^*}b_i\right)+\sum_{i=I^*+2}^{N}(d_i-2)\beta_i.\end{aligned} \tag{2.4.12}$$

To see that (2.4.12) is actually nonnegative for all $x \in X$ we have to distinguish cases for α^*.

Case 1: $\alpha^* = \sum_{i=1}^{l} b_i$

Then $I^* = l$ and (2.4.12) reads

$$\begin{aligned}&\nabla F(x^*)^T(x-x^*)\\&=\sum_{i=1}^{l}(d_i-2)(\beta_i-b_i)+(d_{l+1}-2)\left(\beta_{l+1}-\sum_{i=1}^{l}b_i+\sum_{i=1}^{l}b_i\right)+\sum_{i=l+2}^{N}(d_i-2)\beta_i\\&=\sum_{i=1}^{l}\underbrace{(d_i-2)}_{\le 0}\underbrace{(\beta_i-b_i)}_{\le 0}+\sum_{i=l+1}^{N}\underbrace{(d_i-2)}_{>0}\beta_i\ge 0\qquad\forall(\alpha,\beta)\in X.\end{aligned}$$

In particular, x^* solves the variational inequality (2.4.11) and is thus the global minimum of F.

Case 2: $\alpha^* = c$

In this case we have $I^* \le l$ but we can assume without loss of generality that $I^*+1 \le l$ as otherwise $I^* = l$ and thus $c = \sum_{i=1}^{l} b_i$ which has already been covered in the first case. Hence, (2.4.12) reads

$$\begin{aligned}&\nabla F(x^*)^T(x-x^*)\\&=\sum_{i=1}^{I^*}(d_i-2)(\beta_i-d_i)+(d_{I^*+1}-2)\left(\beta_{I^*+1}-c+\sum_{i=1}^{I^*}b_i\right)+\sum_{i=I^*+2}^{N}(d_i-2)\beta_i\\&=\sum_{i=1}^{I^*}(d_i-2)(\beta_i-b_i)+(d_{I^*+1}-2)\left(\beta_{I^*+1}-c+\sum_{i=1}^{I^*}b_i\right)\\&\quad+\sum_{i=I^*+2}^{l}(d_i-2)\beta_i+\sum_{i=l+1}^{N}(d_i-2)\beta_i.\end{aligned} \tag{2.4.13}$$

In this case bounding the right-hand side from below is not as easy as the β_i are linked together via the constraint $\sum_{i=1}^{N} \beta_i = \alpha$. Nevertheless, the last term is non-negative for all $\beta \in X$, whereas all the terms in the first three terms of the sum are monotonically decreasing in β_i and β_{I^*+1}, respectively. Consequently, we set $\beta_i = 0$ for $i > l$ which monotonically increases the values of the remaining β_i and thus

$$
\begin{aligned}
&\nabla F(x^*)^T(x - x^*) \\
\geq &\sum_{i=1}^{I^*}(d_i - 2)(\beta_i - b_i) + (d_{I^*+1} - 2)\left(\beta_{I^*+1} - c + \sum_{i=1}^{I^*} b_i\right) + \sum_{i=I^*+2}^{l}(d_i - 2)\beta_i \\
= &\sum_{i=1}^{I^*} d_i(\beta_i - b_i) - 2\sum_{i=1}^{I^*}\beta_i + 2\sum_{i=1}^{I^*} b_i + d_{I^*+1}\left(\beta_{I^*+1} + \sum_{i=1}^{I^*} b_i - c\right) \\
&- 2\beta_{I^*+1} + 2c - 2\sum_{i=1}^{I^*} b_i + \sum_{i=I^*+2}^{l} d_i\beta_i - 2\sum_{i=I^*+2}^{l}\beta_i \\
= &\sum_{i=1}^{I^*} d_i(\beta_i - b_i) - 2\sum_{i=1}^{l}\beta_i + d_{I^*+1}\left(\beta_{I^*+1} + \sum_{i=1}^{I^*} b_i - c\right) + 2c + \sum_{i=I^*+2}^{l} d_i\beta_i \\
= &\sum_{i=1}^{I^*} d_i(\beta_i - b_i) + 2(c - \alpha) + d_{I^*+1}\left(\beta_{I^*+1} + \sum_{i=1}^{I^*} b_i - c\right) + \sum_{i=I^*+2}^{l} d_i\beta_i,
\end{aligned}
$$

where we used the improved constraint $\sum_{i=1}^{l} \beta_i = \alpha$. Using the constraint again and rearranging terms gives

$$
\begin{aligned}
&\nabla F(x^*)^T(x - x^*) \\
\geq &\sum_{i=1}^{I^*} d_i(\beta_i - b_i) + 2(c - \alpha) + d_{I^*+1}\left(\alpha - \sum_{i=1}^{I^*}\beta_i - \sum_{i=I^*+2}^{l}\beta_i + \sum_{i=1}^{I^*} b_i - c\right) + \sum_{i=I^*+2}^{l} d_i\beta_i \\
= &\sum_{i=1}^{I^*} d_i(\beta_i - b_i) + 2(c - \alpha) + d_{I^*+1}(\alpha - c) + d_{I^*+1}\sum_{i=1}^{I^*}(b_i - \beta_i) + \sum_{i=I^*+2}^{l}(d_i - d_{I^*+1})\beta_i \\
= &\sum_{i=1}^{I^*}\underbrace{(d_{I^*+1} - d_i)}_{\geq 0}\underbrace{(b_i - \beta_i)}_{\geq 0} + \underbrace{(c - \alpha)}_{\geq 0}\underbrace{(2 - d_{I^*+1})}_{\geq 0} + \sum_{i=I^*+2}^{l}\underbrace{(d_i - d_{I^*+1})}_{\geq 0}\beta_i \geq 0
\end{aligned}
\tag{2.4.14}
$$

for all $x \in X$. Note that we used once more that $I^* + 1 \leq l$, so that $d_{I^*+1} \leq 2$. From (2.4.14) we conclude that x^* solves the variational inequality (2.4.11) also in the case $\alpha^* = c$ and x^* is thus the global minimum of F. □

Remark 2.4.10. Now we give a heuristic for the optimal choice (α^*, β^*) given in equation (2.4.10). In a first step, we fix the amount of mass α that we want to transport and look for the optimal weights β given α. As $\beta_i = 0$ for all $i = 1, ..., n$ if $\alpha = 0$, we will only consider $\alpha > 0$. According to (2.4.9) we have the estimate

$$\rho_F(\mu, \nu) \leq c + \sum_{i=1}^{N} b_i - 2\alpha + \sum_{i=1}^{N} \beta_i d_i. \tag{2.4.15}$$

As the Wasserstein distance scales with the transported distance, it is clearly optimal to transport mass as shortly as possible. Since the d_i are ordered increasingly, this means that we prioritize lower indices over higher ones when assigning mass for transportation. To this end, let

$$I_\alpha := \max \left\{ I \in \{0, ..., N\} \mid \sum_{i=1}^{I} b_i \leq \alpha \right\},$$

i.e. I_α denotes the index up to which we can assign the maximal value to β_i, i.e. $\beta_i = b_i$, as there is still sufficient mass left assigned for transportation. The remaining mass $\alpha - \sum_{i=1}^{I_\alpha} b_i$ is then delegated to the next entry $I_\alpha + 1$. All other entries are set to zero, so that this scheme yields the following weight vector $\beta^* = \beta^*(\alpha)$:

$$\beta_i^* = \begin{cases} b_i, & i \leq I_\alpha, \\ \alpha - \sum_{i=1}^{I_\alpha} b_i, & i = I_\alpha + 1, \\ 0, & \text{else.} \end{cases} \tag{2.4.16}$$

We note that distributing the mass in any other way by choosing different β_i can not yield a better overall transportation cost $W_1(\tilde{\mu}_\alpha, \nu_{\alpha,\beta})$ as we would potentially transport more mass to locations further away at the expense of nearer locations. However, equally efficient transport plans might be possible if there are points with the same distance to x_0 so that mass transportation is indifferent between those locations.

Now we are left to find the optimal choice of α which minimizes (2.4.15). So we define

$$\begin{aligned}\tilde{F}(\alpha) :=& F(\alpha,\beta^*) = c + \sum_{i=1}^N b_i - 2\alpha + \sum_{i=1}^N \beta_i^* d_i \\ =& c + \sum_{i=1}^N b_i - 2\alpha + \sum_{i=1}^{I_\alpha} b_i d_i + \left(\alpha - \sum_{i=1}^{I_\alpha} b_i\right) d_{I_\alpha+1} \\ =& \underbrace{c + \sum_{i=1}^N b_i}_{=:K} + \sum_{i=1}^{I_\alpha} b_i(d_i-2) + (d_{I_\alpha+1}-2)\left(\alpha - \sum_{i=1}^{I_\alpha} b_i\right)\end{aligned} \tag{2.4.17}$$

and want to minimize with respect to α.
We claim that α is actually bounded by $\sum_{i=1}^l b_i$, so that

$$\alpha \le \min\left\{c, \sum_{i=1}^l b_i\right\} \qquad \text{and} \qquad I_\alpha = \max\left\{I \in \{0,...,l\} \mid \sum_{i=1}^I b_i \le \alpha\right\}. \tag{2.4.18}$$

To this end, first consider the case that $\alpha > \sum_{i=1}^l b_i$. Then clearly $I_\alpha \ge l$. If $I_\alpha = l$, then we see

$$\begin{aligned}\tilde{F}_l := \tilde{F}(\alpha) =& K + \sum_{i=1}^l b_i(d_i-2) + \underbrace{(d_{l+1}-2)}_{>0}\underbrace{\left(\alpha - \sum_{i=1}^l b_i\right)}_{\ge 0} \\ >& K + \sum_{i=1}^l b_i(d_i-2) = F\left(\sum_{i=1}^l b_i\right),\end{aligned} \tag{2.4.19}$$

so that choosing $\alpha = \sum_{i=1}^l b_i$ leads to a smaller value. On the other hand, if $I_\alpha \ge l+1$, then due to the monotonicity of the d_i

$$\begin{aligned}\tilde{F}(\alpha) =& K + \sum_{i=1}^l b_i(d_i-2) + \sum_{i=l+1}^{I_\alpha} b_i(d_i-2) + (d_{I_\alpha+1}-2)\left(\alpha - \sum_{i=1}^{I_\alpha} b_i\right) \\ \ge& K + \sum_{i=1}^l b_i(d_i-2) + \sum_{i=l+1}^{I_\alpha} b_i(d_i-2) + (d_{l+1}-2)\left(\alpha - \sum_{i=1}^{I_\alpha} b_i\right) \\ =& K + \sum_{i=1}^l b_i(d_i-2) + (d_{l+1}-2)\left(\alpha - \sum_{i=1}^l b_i\right) + \sum_{i=l+1}^{I_\alpha} b_i(d_i-2) - \sum_{i=l+1}^{I_\alpha}(d_{l+1}-2)b_i \\ =& \tilde{F}_l + \sum_{i=l+1}^{I_\alpha} b_i \underbrace{(d_i - d_{l+1})}_{\ge 0} \ge \tilde{F}_l.\end{aligned}$$

As we have seen in (2.4.19) $\tilde{F}_l$ can not be minimal and therefore neither any α with $I_\alpha \geq l+1$. We conclude that indeed (2.4.18) holds.

We come back to minimizing $\tilde{F}$ in (2.4.17) and claim that $\tilde{F}$ is monotonically decreasing for $\alpha \in \left[0, \min\left\{c, \sum_{i=1}^l b_i\right\}\right]$ which will conclude this proof. Indeed, if $\tilde{F}$ is montonically decreasing, it takes its minimum at the right hand side of the interval and thus $\alpha^* = \min\left\{c, \sum_{i=1}^l b_i\right\}$.

For the proof of monotonicity, we first note that without loss of generality we can assume $I_\alpha + 1 \leq l$, so that in the last term of (2.4.17) we have $d_{I_\alpha+1} \leq 2$. Indeed, $I_\alpha + 1 > l$ implies $I_\alpha > l-1$, or rather $I_\alpha = l$ and thus $\alpha = \sum_{i=1}^l b_i$ with the third term vanishing completely in this case.
Now we turn to the monotonicity of $\tilde{F}$. Let $\alpha_1 \leq \alpha_2 \leq \min\left\{c, \sum_{i=1}^l b_i\right\}$. If $I_{\alpha_1} = I_{\alpha_2} =: I$, then

$$\tilde{F}(\alpha_1) - \tilde{F}(\alpha_2) = \underbrace{(d_{I+1} - 2)}_{\leq 0}(\alpha_1 - \alpha_2) \geq 0,$$

so $\tilde{F}$ is indeed monotonically decreasing in this case. If $I_{\alpha_1} < I_{\alpha_2} \leq l$, then

$$\begin{aligned}
&\tilde{F}(\alpha_1) - \tilde{F}(\alpha_2) \\
=&K + \sum_{i=1}^{I_{\alpha_1}} b_i(d_i - 2) + (d_{I_{\alpha_1}+1} - 2)\left(\alpha_1 - \sum_{i=1}^{I_{\alpha_1}} b_i\right) \\
&- K - \sum_{i=1}^{I_{\alpha_2}} b_i(d_i - 2) - (d_{I_{\alpha_2}+1} - 2)\left(\alpha_2 - \sum_{i=1}^{I_{\alpha_1}} b_i\right) \\
=& - \sum_{i=I_{\alpha_1}+1}^{I_{\alpha_2}} b_i d_i + (d_{I_{\alpha_1}+1} - 2)\alpha_1 - d_{I_{\alpha_1}+1}\sum_{i=1}^{I_{\alpha_1}} b_i - (d_{I_{\alpha_2}+1} - 2)\alpha_2 + d_{I_{\alpha_2}+1}\sum_{i=1}^{I_{\alpha_2}} b_i.
\end{aligned}$$

We rearrange and both add and substract auxiliary terms which leads to

$$
\begin{aligned}
&\tilde{F}(\alpha_1) - \tilde{F}(\alpha_2)\\
=&d_{I_{\alpha_2}+1}\sum_{i=1}^{I_{\alpha_2}} b_i - d_{I_{\alpha_1}+1}\sum_{i=1}^{I_{\alpha_2}} b_i + d_{I_{\alpha_1}+1}\sum_{i=1}^{I_{\alpha_2}} b_i - \sum_{i=I_{\alpha_1}+1}^{I_{\alpha_2}} b_i d_i - d_{I_{\alpha_1}+1}\sum_{i=1}^{I_{\alpha_1}} b_i\\
&+ (d_{I_{\alpha_1}+1} - d_{I_{\alpha_2}+1})\alpha_1 + d_{I_{\alpha_2}+1}(\alpha_1 - \alpha_2) + 2(\alpha_2 - \alpha_1)\\
=&(d_{I_{\alpha_2}+1} - d_{I_{\alpha_1}+1})\left(\sum_{i=1}^{I_{\alpha_2}} b_i - \alpha_1\right) + d_{I_{\alpha_1}+1}\sum_{i=I_{\alpha_1}+1}^{I_{\alpha_2}} b_i - \sum_{i=I_{\alpha_1}+1}^{I_{\alpha_2}} b_i d_i\\
&+ (\alpha_2 - \alpha_1)(2 - d_{I_{\alpha_2}+1})\\
=&(d_{I_{\alpha_2}+1} - d_{I_{\alpha_1}+1})\left(\sum_{i=1}^{I_{\alpha_1}} b_i - \alpha_1\right) + (d_{I_{\alpha_2}+1} - d_{I_{\alpha_1}+1})\sum_{i=I_{\alpha_1}+1}^{I_{\alpha_2}} b_i\\
&+ \sum_{i=I_{\alpha_1}+1}^{I_{\alpha_2}} b_i(d_{I_{\alpha_1}+1} - d_i) + (\alpha_2 - \alpha_1)(2 - d_{I_{\alpha_2}+1})\\
=&(d_{I_{\alpha_2}+1} - d_{I_{\alpha_1}+1})\left(\sum_{i=1}^{I_{\alpha_1}} b_i - \alpha_1\right) + \sum_{i=I_{\alpha_1}+1}^{I_{\alpha_2}} b_i(d_{I_{\alpha_2}+1} - d_i) + (\alpha_2 - \alpha_1)(2 - d_{I_{\alpha_2}+1}).
\end{aligned}
$$

Now using that $\sum_{i=1}^{I_{\alpha_1}+1} b_i - \alpha_1 > 0$ by definition of I_{α_1}, we conclude that $\tilde{F}$ is also monotonically decreasing in this case

$$
\begin{aligned}
&\tilde{F}(\alpha_1) - \tilde{F}(\alpha_2)\\
=&\underbrace{(d_{I_{\alpha_2}+1} - d_{I_{\alpha_1}+1})}_{\geq 0}\underbrace{\left(\sum_{i=1}^{I_{\alpha_1}+1} b_i - \alpha_1\right)}_{>0} + \underbrace{\sum_{i=I_{\alpha_1}+2}^{I_{\alpha_2}} b_i(d_{I_{\alpha_2}+1} - d_i)}_{\geq 0} + \underbrace{(\alpha_2 - \alpha_1)(2 - d_{I_{\alpha_2}+1})}_{\geq 0}\\
\geq&0.
\end{aligned}
$$

It is important to note that this approach just provides a heuristic for the global minimum (α^*, β^*) and that a proof for the optimality is still necessary. In this remark we computed

$$\min_{\alpha}\min_{\beta} F(\alpha,\beta) \qquad \text{instead of} \qquad \min_{\alpha,\beta} F(\alpha,\beta)$$

under the constraint $\sum_{i=1}^{N}\beta_i = \alpha$. In general, the second term is smaller than the first.

2.5 Computing the flat metric

So far, our studies concerning the flat metric have been purely theoretical. However, if we want to capitalize on the similarity to the Wasserstein distances for applications in unbalanced optimal transport tasks, we need a way to explicitly calculate the flat distance between measures $\mu, \nu \in \mathcal{M}^+(\mathbb{R}^d)$. As we have seen in Proposition 2.4.8, closed analytical expressions are complicated even for Dirac measures, and are in general unattainable. Thus, instead we apply machine learning to calculate the flat distance. The method is based on the work by Cem Anil, James Lucas and Roger Grosse [9] which uses the Kantorovich-Rubinstein duality (2.4.1) to compute the Wasserstein distance W_1. Since the (dual) formulation (2.2.3) of the flat metric is of a similar structure, we were able to adjust the concepts to our setting [135].
We summarize the most important points of the paper as a thorough introduction to the topic and applied methods would go beyond the scope of this thesis. The complete code as well as examples and visualisation tools can be found at `https://github.com/hs42/flat_metric`. The main idea is to train a neural network f_Θ to find the optimal test function realizing the flat distance between μ and ν in (2.2.3). The network is given by a **multi-layer perceptron** with two fully connected hidden layers having 64 neurons each and the Adam optimizer [92]. We deliberately choose a shallow network architecture as it provides sufficiently good results whereas moving to larger networks results in instabilities or even failures during training due to limited training data. Indeed, in our setting, training data for the network is given by the support points of the measures μ, ν which are usually sparse for the (mostly discrete) measures appearing applications. Despite the shallow architecture, the Universal Approximation Theorem proven in [9] guarantees that the entire space $BL(\mathbb{R}^d)$ can be accessed via the network if the architectural constraints are chosen appropriately, so that we can expect meaningful results.

For two given measures $\mu, \nu \in \mathcal{M}^+(\mathbb{R}^d)$ we make the ansatz $\psi = f_\Theta$ and model the optimal bounded Lipschitz test function by the neural network. To ensure that f_Θ is indeed admissible to the optimisation problem (2.2.3), i.e. that it is a bounded Lipschitz function with $\|\cdot\|_{BL}$ norm bounded by 1, we use a mixed approach of architectural constraints and regularisation. More precisely, we adopt the architectural approach introduced in [9] to guarantee Lipschitz continuity whereas we use

additional regularisational constraints to enforce boundedness of f_Θ.
Architectural constraints: The method introduced in [9] to calculate the Wasserstein distance W_1 takes advantage of the fact that Lipschitz continuity is closed under compositions, so that it is sufficient to control the Lipschitz constant of each individual layer and each activation function to achieve Lipschitz continuity of the output. In the paper [9], the authors Anil, Lucas and Grosse normalise each layer A_i using Björck orthonormalisation [17] and use the nonlinear 1-Lipschitz shuffling operator **GroupSort** [26] as activation function. The Björck orthonormalisation ensures that the linear transformation induced by layer A_i is in fact isometric, thus strictly enforcing $|A_i|_{\mathbf{Lip}} = 1$. However, in contrast to the Wasserstein distance where the optimal test function will always be exactly 1-Lipschitz, the optimal test function for the flat metric often has a smaller Lipschitz constant, rendering the Björck orthonormalisation too restrictive for our purposes. Thus, in our implementation we apply a **spectral normalisation** leading to $\|A_i\|_2 = 1$ instead. This guarantees that the largest singular value is exactly 1 but there may be other eigenspaces with smaller absolute singular values. Consequently, layer A_i is not necessarily 1-Lipschitz in every direction but only satisfies $|A_i|_{\mathbf{Lip}} \leq 1$. In the cases we considered, both Björck and spectral normalisation provided similar results, but from a theoretical perspective spectral normalisation does not preserve the gradient norm, unlike Björck orthonormalisation, which potentially leads to diminishing gradient norms of the network during backpropagation and thus to a slower convergence of the network, see [9, B.2]. We remark that as a side effect, Lipschitz constrained neural networks are proven to be adversarially robust which means that the change in output under small adversarial perturbations is bounded [155].
Regularisation constraints: Since the networks architecture ensures Lipschitz continuity of the output, we need to introduce a loss term that both renders f_Θ admissible the optimisation problem (2.2.3) as well as yields the optimal value for the flat metric. To this end, our **total loss term** $\mathcal{L}$ consists of two parts

$$\mathcal{L} := \mathcal{L}_m + \lambda \mathcal{L}_b. \tag{2.5.1}$$

Such an approach of having one loss term for the problem and one for the admissibility is commonly employed, e.g in the implementation of Wasserstein gradient-penalty adversarial networks [77]. The **metric loss** term $\mathcal{L}_m$ corresponds to minimising the negative of (2.2.3) and is given by

$$\mathcal{L}_m := -\int_{\mathbb{R}^d} f_\Theta(x)\,\mathrm{d}\mu(x) + \int_{\mathbb{R}^d} f_\Theta(x)\,\mathrm{d}\nu(x). \tag{2.5.2}$$

Thus, after training $\mathcal{L}_m$ approximately yields the negative value of the flat distance $\rho_F(\mu,\nu)$. Additionally, we have the **bound loss term** which serves as a penalty term to bound f_Θ

$$\mathcal{L}_b\left(\frac{1}{\|\mu\|_{TV}}\langle h_\mu, h_\mu\rangle + \frac{1}{\|\nu\|_{TV}}\langle h_\nu, h_\nu\rangle\right). \tag{2.5.3}$$

Here, $h_\kappa := \max_{x\sim\kappa}(|f_\Theta(x)| - M, 0)$, where M denotes the desired upper bound of $\|f_\Theta\|_\infty$; in our setting $M = 1$. In contrast to the obvious choice of simply considering the maximal value $\|f_\Theta\|_\infty$, we choose this approach to reduce the effect of outliers in the data and thus to simplify training of the network.
The functions h_κ encode in which areas f_Θ deviates from its target bound evaluated each on the support points of $\kappa = \mu$ and $\kappa = \nu$ respectively. If such a deflection $|f_\Theta| > M$ occurs, the corresponding h_κ will have non-vanishing values and h_κ serves as a penalty. The penalties are then accumulated over the whole space by the inner product $\langle\cdot,\cdot\rangle$ which thus measures how much f_Θ violates the bound when evaluated with respect to μ and ν respectively. The loss term should not favour measures with larger total masses, so that each contribution is normalised by its respective total variation ensuring that the penalty terms remain invariant under scaling of the total mass. This has the useful side effect that it is not important for our implementation whether the same empiric distribution is represented using 100 or 1000 data points. Finally, the penalty contributions with respect to μ and ν are combined to give the overall penalty $\mathcal{L}_b$. In practice, strictly enforcing the ideal bound of a vanishing $\mathcal{L}_b$ is not possible and hence we strive for small values of the loss instead.

Note that the two loss contributions $\mathcal{L}_m$ and $\mathcal{L}_b$ of $\mathcal{L}$ in (2.5.1) are balanced by an enforcing parameter $\lambda = \lambda(t)$ depending on elapsed training time t. Specifically, λ is chosen *adaptively* so that each freshly trained network is approximately bound by the same constant $\|f_\Theta\|_\infty \leq M$ while simultaneously having comparable relative loss contributions of $\mathcal{L}_m$ and $\mathcal{L}_b$, regardless of the input distributions. This is particularly important for our setting as we want to establish pairwise comparisons of neural networks which have been trained independently and/or on different data sets, so that the output of the network should be ordinal. This occurs for example, when computing pairwise distances between subdistributions. Without dynamically balancing the loss terms, the resulting f_Θ will adhere more or less strict to the $\|\cdot\|_\infty$ bound, depending on the currently dominating loss term, leading to biased results. Notably, different networks would solve different optimisation problems (2.2.3) yielding their actual outcomes to be incomparable to each other.

We performed several experiments varying data dimensions and network parameters to test our method against available ground truth (see Proposition 2.4.8). The results were satisfying and showed that we were able to capture the important aspects of the flat metric. For more information, we refer to [135]. To conclude this summary, we present an exemplary experiment which also demonstrates fields of applications for the method.
As a first step, we simulate high dimensional single-cell (sc) transcriptomics data using the R-software package *Splatter* which has been developed by Zappia et al. [166] to generate simulated scRNA sequencing count data of differentiation trajectories or of populations with one or multiple cell types. In this setting, analytical ground truth is unavailable, but we still have the possibility to monitor qualitative changes of the implementation via appropriate parameter choices in the *Splatter* framework. In particular, we modelled 10 000 cells, distributed over five different cell groups with varying sample sizes and genetic expression profiles, i.e. locations in gene space. After preprocessing and reducing the generated data to the first 50 principal dimensions, we determined the flat distances between the individual groups, see Table 2.1. For comparison, we computed the corresponding Wasserstein distances of the normalised distributions as well. For visualisation, we further reduced the data to 2 dimensions and generated a t-SNE plot, see Figure 2.3.

Table 2.1: Flat distances (first entry of each cell) between the clusters for dimension 50. For comparison the respective Wasserstein distances using the same net architecture are displayed as well (second entry of each cell)

	Group 1	Group 2	Group 3	Group 4	Group 5
Group 1	(0.00, 0.00)	(3.15, 0.35)	(2.17, 6.77)	(5.15, 6.85)	(2.25, 9.75)
Group 2	(3.16, 0.35)	(0.00, 0.00)	(5.16, 6.76)	(2.16, 6.85)	(4.79, 9.76)
Group 3	(2.17, 6.77)	(5.15, 6.76)	(0.00, 0.00)	(5.21, 9.51)	(2.26, 11.86)
Group 4	(5.15, 6.85)	(2.16, 6.68)	(5.20, 9.51)	(0.00, 0.00)	(4.83, 12.05)
Group 5	(2.25, 9.75)	(4.79, 9.76)	(2.27, 11.86)	(4.84, 12.05)	(0.00, 0.00)

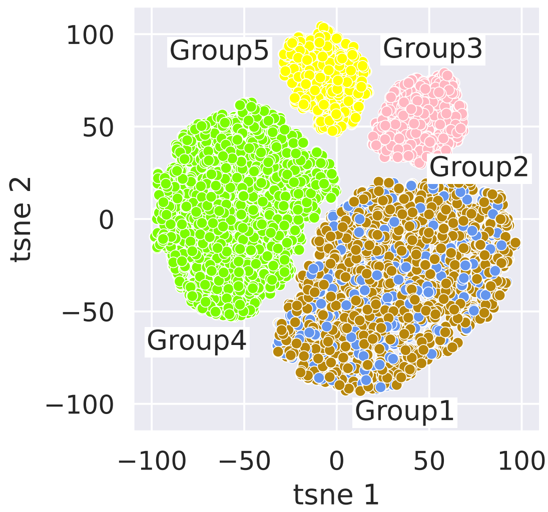

Figure 2.3: 2D t-SNE plot of mRNA counts for 5 distributions generated by *Splatter*.

Table 2.1 shows the systematic differences between the flat metric and the Wasserstein distance. As the latter is insensitive to population size, distributions 1 (blue) and 2 (brown) are nearly identical in Wasserstein space whereas they are clearly distinguishable with respect to the flat metric due to the large mass difference. The mass imbalances thus significantly influence the neighbouring relation of the groups. So, in situations in which differences in cluster sizes are not only an effect of sampling but rather play a relevant role for the underlying question, we highly recommend using a method for unnormalised data distributions such as our flat metric.

3 PDE models on the Euclidean space

In this chapter we will investigate a generic structured population model in an Euclidean setting. Starting from a classical PDE model, we derive a weak formulation in measures using the Theorem of Radon-Nikodým. Sections 3.3 and 3.4 deal with existence and uniqueness of solutions to a linear version of the model, whereas the general nonlinear model will be considered in Section 3.5.

3.1 From a PDE model to a weak formulation

In the first part of this chapter we treat a linear structured population model on $[0,T] \times \mathbb{R}^d$ of the form

$$\begin{cases} \partial_t \mu_t + \nabla_x \cdot (b(t,x)\mu_t) & = c(t,x)\mu_t + \int_{\mathbb{R}^d} \eta(t,x)(y)\,\mathrm{d}\mu_t(y) + N(t), \\ \mu_0 & = \nu, \end{cases} \tag{3.1.1}$$

where $\nu \in \mathcal{M}^+(\mathbb{R}^d)$ and

$$\begin{aligned} b &: [0,T] \times \mathbb{R}^d \to \mathbb{R}^d, & c &: [0,T] \times \mathbb{R}^d \to \mathbb{R}, \\ \eta &: [0,T] \times \mathbb{R}^d \to \mathcal{M}^+(\mathbb{R}^d), & N &: [0,T] \to \mathcal{M}^+(\mathbb{R}^d). \end{aligned}$$

Function c represents a **growth term**, η can be interpreted as **spread of heterogeneity** (such as a **mutation kernel**), N is a **state-independent influx** and b denotes the **vector field** which is responsible for the transformation dynamic of the individual states.

Unless individual results assume a different regularity, we assume the following for our model functions.

Assumptions 3.1.1. The model functions c, η and X satisfy

(i) $b \in C^0([0,T]; BL(\mathbb{R}^d; \mathbb{R}^d))$,

(ii) $c \in C^0([0,T]; BL(\mathbb{R}^d))$,

(iii) $\eta \in C^0\left([0,T]; BL(\mathbb{R}^d; \mathcal{M}^+(\mathbb{R}^d))\right)$,

(iv) $N \in C^0([0,T]; \mathcal{M}^+(\mathbb{R}^d))$.

Notation 3.1.2. *To explain the technical assumptions on the model functions, we elaborate the most complicated assumption (iii) from a computational point of view. For fixed $t \in [0,T]$ and $x \in \mathbb{R}^d$, $\eta(t,x)$ is a measure in $\mathcal{M}^+(\mathbb{R}^d)$, whereas for fixed t the map $x \mapsto \eta(t,x) \in \mathcal{M}^+(\mathbb{R}^d)$ is bounded and Lipschitz continuous, i.e.*

$$\|\eta(t,\cdot)\|_{BL(\mathbb{R}^d;\mathcal{M}^+)} := \max\left\{\sup_{x\in\mathbb{R}^d} \|\eta(t,x)\|_{BL^*}, |\eta(t,\cdot)|_{\mathbf{Lip}}\right\} < \infty,$$

where the Lipschitz constant is given by

$$|\eta(t,\cdot)|_{\mathbf{Lip}} = \sup_{x_1\neq x_2} \frac{\|\eta(t,x_1) - \eta(t,x_2)\|_{BL^*}}{|x_1 - x_2|}.$$

Finally, $t \mapsto \|\eta(t,\cdot)\|_{BL(\mathbb{R}^d;\mathcal{M}^+)}$ is continuous so that

$$\|\eta\|_\infty := \sup_{t\in[0,T]} \|\eta(t,\cdot)\|_{BL(\mathbb{R}^d;\mathcal{M}^+)} < \infty.$$

In a similar vein, we introduce the following for abbreviation

$$\begin{aligned}
\|b\|_\infty &:= \sup_{t\in[0,T]} \|b(t,\cdot)\|_{BL} < \infty, \\
\|c\|_\infty &:= \sup_{t\in[0,T]} \|c(t,\cdot)\|_{BL} < \infty, \\
\|N\|_\infty &:= \sup_{t\in[0,T]} \|N(t)\|_{BL^*} < \infty.
\end{aligned}$$

We start with deriving the weak formulation of (3.1.1). To this end we first assume higher regularity, i.e. we don't consider measures but rather the problem

$$\begin{cases} \partial_t u_t(x) + \nabla_x \cdot (b(t,x)u_t(x)) = c(t,x)u_t(x) + \tilde{n}(t,x)u_t(x) + \tilde{N}(t) & \in [0,T]\times\mathbb{R}^d \\ u_0(x) = h(x), \end{cases} \tag{3.1.2}$$

where $u_t(x) := u(t,x)$ and $u \in C^1([0,T]\times\mathbb{R}^d)$. Multiplying (3.1.2) with a test function $\varphi \in BL([0,T]\times\mathbb{R}^d) \cap C^1([0,T]\times\mathbb{R}^d)$ and integrating over the whole domain yields

$$\begin{aligned}
&\int_0^T \int_{\mathbb{R}^d} \partial_t u_t(x)\varphi(t,x) + \nabla_x \cdot (b(t,x)u_t(x))\varphi(t,x)\,\mathrm{d}x\,\mathrm{d}t \\
&\qquad = \int_0^T \int_{\mathbb{R}^d} c(t,x)u_t(x)\varphi(t,x) + \tilde{n}(t,x)u_t(x)\varphi(t,x) + \tilde{N}(t)\varphi(t,x)\,\mathrm{d}x\,\mathrm{d}t.
\end{aligned} \tag{3.1.3}$$

For the first term on the left hand side of (3.1.3) we apply Fubini's Theorem twice and integrate by parts which gives

$$\begin{aligned}
&\int_0^T \int_{\mathbb{R}^d} \partial_t u_t(x)\varphi(t,x)\,\mathrm{d}x\,\mathrm{d}t = \int_{\mathbb{R}^d} \int_0^T \partial_t u_t(x)\varphi(t,x)\,\mathrm{d}t\,\mathrm{d}x \\
&= -\int_{\mathbb{R}^d} \int_0^T u_t(x)\partial_t\varphi(t,x)\,\mathrm{d}t\,\mathrm{d}x + \int_{\mathbb{R}^d} [u_t(x)\varphi(t,x)]_0^T\,\mathrm{d}x \\
&= -\int_0^T \int_{\mathbb{R}^d} u_t(x)\partial_t\varphi(t,x)\,\mathrm{d}x\,\mathrm{d}t + \int_{\mathbb{R}^d} u_T(x)\varphi(T,x) - u_0(x)\varphi(0,x)\,\mathrm{d}x.
\end{aligned} \tag{3.1.4}$$

The second term on the left hand side of (3.1.3) can be rewritten with Green's identity

$$\int_0^T \int_{\mathbb{R}^d} \nabla_x \cdot (b(t,x)u_t(x))\varphi(t,x)\,\mathrm{d}x\,\mathrm{d}t = -\int_0^T \int_{\mathbb{R}^d} b(t,x)u_t(x) \cdot \nabla_x\varphi(t,x)\,\mathrm{d}x\,\mathrm{d}t. \tag{3.1.5}$$

Plugging both (3.1.4) and (3.1.5) into (3.1.3) and rearranging yields

$$\begin{aligned}
&\int_{\mathbb{R}^d} u_T(x)\varphi(T,x) - u_0(x)\varphi(0,x)\,\mathrm{d}x \\
&= \int_0^T \int_{\mathbb{R}^d} u_t(x)\partial_t\varphi(t,x) + b(t,x)u_t(x) \cdot \nabla_x\varphi(t,x) + c(t,x)u_t(x)\varphi(t,x)\,\mathrm{d}x\,\mathrm{d}t \\
&\quad + \int_0^T \int_{\mathbb{R}^d} \tilde{n}(t,x)u_t(x)\varphi(t,x) + \tilde{N}(t)\varphi(t,x)\,\mathrm{d}x\,\mathrm{d}t.
\end{aligned} \tag{3.1.6}$$

In a second step, we assume that u_t is actually the density of a measure μ_t with respect to the Lebesgue measure λ^d according to Radon-Nikodým Theorem [7, Theorem 1.28]. In particular, we have

$$u_t(x) = D_{\lambda^d}\mu_t(x).$$

Furthermore, the model functions $\tilde{n}$ and $\tilde{N}$ are supposed to be densities as well, i.e.

$$\tilde{n}(t,x) = D_{\mu_t}\nu_t(x) \qquad \text{and } \tilde{N}(t) = D_{\lambda^d}N(t),$$

where $\nu_t(y) = \left[\int_{\mathbb{R}^d} \eta(t,x)\,\mathrm{d}\mu_t(x)\right](y)$. Note that it is necessary to apply the Radon-Nikodým Theorem for Banach spaces to treat $\tilde{n}(t,x)$ (see [126]).
Using basic properties of the Radon-Nikodým density (see [44, F.22]), we can rewrite the terms in (3.1.6) as follows

$$\begin{aligned}
&\bullet \int_{\mathbb{R}^d} u_T(x)\varphi(T,x) - u_0(x)\varphi(0,x)\,\mathrm{d}x = \int_{\mathbb{R}^d} \varphi(T,x)\,\mathrm{d}\mu_T(x) - \int_{\mathbb{R}^d} \varphi(0,x)\,\mathrm{d}\mu_0(x),\\
&\bullet \int_0^T \int_{\mathbb{R}^d} u_t(x)\partial_t\varphi(t,x)\,\mathrm{d}x\,\mathrm{d}t = \int_0^T \int_{\mathbb{R}^d} \partial_t\varphi(t,x)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t,\\
&\bullet \int_0^T \int_{\mathbb{R}^d} b(t,x)u_t(x)\cdot\nabla_x\varphi(t,x)\,\mathrm{d}x\,\mathrm{d}t = \int_0^T \int_{\mathbb{R}^d} b(t,x)\cdot\nabla_x\varphi(t,x)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t,\\
&\bullet \int_0^T \int_{\mathbb{R}^d} \tilde{N}(t)\varphi(t,x)\,\mathrm{d}x\,\mathrm{d}t = \int_0^T \int_{\mathbb{R}^d} \varphi(t,x)\,\mathrm{d}N(t)(x)\,\mathrm{d}t.
\end{aligned} \tag{3.1.7}$$

It remains to rewrite the term involving $\tilde{n}$. As linear operations commute with the Bochner integral (see [164, V.5. Corollary 2]), we see that

$$\begin{aligned}
&\int_0^T \int_{\mathbb{R}^d} \tilde{n}(t,x)u_t(x)\varphi(t,x)\,\mathrm{d}x\,\mathrm{d}t = \int_0^T \int_{\mathbb{R}^d} \tilde{n}(t,x)\varphi(t,x)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t\\
&= \int_0^T \int_{\mathbb{R}^d} \varphi(t,x)\,\mathrm{d}\nu_t(x)\,\mathrm{d}t = \int_0^T \int_{\mathbb{R}^d} \varphi(t,y)\,\mathrm{d}\nu_t(y)\,\mathrm{d}t\\
&= \int_0^T \int_{\mathbb{R}^d} \varphi(t,y)\,\mathrm{d}\left[\int_{\mathbb{R}^d} \eta(t,x)\,\mathrm{d}\mu_t(x)\right](y)\,\mathrm{d}t\\
&= \int_0^T \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \varphi(t,y)\,\mathrm{d}[\eta(t,x)](y)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t.
\end{aligned} \tag{3.1.8}$$

Plugging (3.1.7) and (3.1.8) into (3.1.6) yields for any test function

$\varphi \in BL([0,T] \times \mathbb{R}^d) \cap C^1([0,T] \times \mathbb{R}^d)$

$$\begin{aligned}
\int_{\mathbb{R}^d} \varphi(T,x)\,\mathrm{d}\mu_T(x) - \int_{\mathbb{R}^d} \varphi(0,x)\,\mathrm{d}\mu_0(x) &= \int_0^T \int_{\mathbb{R}^d} \partial_t\varphi(t,x)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t \\
&+ \int_0^T \int_{\mathbb{R}^d} \big(\nabla_x\varphi(t,x)\cdot b(t,x) + \varphi(t,x)\,c(t,x)\big)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t \\
&+ \int_0^T \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \varphi(t,y)\,\mathrm{d}[\eta(t,x)]\,(y)\right)\mathrm{d}\mu_t(x)\,\mathrm{d}t + \int_0^T \int_{\mathbb{R}^d} \varphi(t,x)\,\mathrm{d}[N(t)](x)\,\mathrm{d}t,
\end{aligned} \tag{3.1.9}$$

which is the **weak formulation** of (3.1.1).

Based on (3.1.9) we introduce the following notion of weak solution to (3.1.1).

Definition 3.1.3. *A family of measures $\mu_\bullet := \{\mu_t\}_{t\in[0,T]} \subset \mathcal{M}^+(\mathbb{R}^d)$ is a **measure solution** to* (3.1.1) *provided $t \mapsto \mu_t$ is narrowly continuous and satisfies* (3.1.9) *for any $\varphi \in BL([0,T] \times \mathbb{R}^d) \cap C^1([0,T] \times \mathbb{R}^d)$.*

We first note that narrow continuity of the measure solution implies that we can replace T with any $s \in [0,T]$ in the weak formulation 3.1.9. More precisely,

Lemma 3.1.4. *Let $s \in [0,T]$ and let $\mu_\bullet$ be a measure solution to* (3.1.9). *Then $\mu_\bullet$ satisfies*

$$\begin{aligned}
\int_{\mathbb{R}^d} \varphi(s,x)\,\mathrm{d}\mu_s(x) - \int_{\mathbb{R}^d} \varphi(0,x)\,\mathrm{d}\mu_0(x) &= \int_0^s \int_{\mathbb{R}^d} \partial_t\varphi(t,x)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t \\
&+ \int_0^s \int_{\mathbb{R}^d} \big(\nabla_x\varphi(t,x)\cdot b(t,x) + \varphi(t,x)\,c(t,x)\big)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t \\
&+ \int_0^s \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \varphi(t,y)\,\mathrm{d}[\eta(t,x)]\,(y)\right)\mathrm{d}\mu_t(x)\,\mathrm{d}t + \int_0^s \int_{\mathbb{R}^d} \varphi(t,x)\,\mathrm{d}[N(t)](x)\,\mathrm{d}t,
\end{aligned}$$

Proof. Let $\varphi \in BL([0,T] \times \mathbb{R}^d) \cap C^1([0,T] \times \mathbb{R}^d)$. For $\varepsilon > 0$ consider a cut-off function $h_\varepsilon : [0,T] \to [0,1]$ which satisfies $h_\varepsilon|_{[0,s]} \equiv 1$ and h_ε decreases linearly to 0 in $[s, s+\varepsilon]$. We plug the λ^{d+1}-a.e. continuously differentiable function $\varphi(t,x)h_\varepsilon(t)$ into the LHS of (3.1.9) and see

$$\int_{\mathbb{R}^d} \varphi(T,x)h_\varepsilon(T)\,\mathrm{d}\mu_T(x) - \int_{\mathbb{R}^d} \varphi(0,x)h_\varepsilon(0)\,\mathrm{d}\mu_0(x) = -\int_{\mathbb{R}^d} \varphi(0,x)\,\mathrm{d}\mu_0(x) \tag{3.1.10}$$

For the RHS we consider the terms separately. We start with the term involving the time derivative

$$
\begin{aligned}
&\int_0^T \int_{\mathbb{R}^d} \partial_t \left(\varphi(t,x) h_\varepsilon(t)\right) \mathrm{d}\mu_t(x)\,\mathrm{d}t \\
=&\int_0^T \int_{\mathbb{R}^d} (\partial_t \varphi(t,x)) h_\varepsilon(t) + \varphi(t,x) \frac{\mathrm{d}}{\mathrm{d}t} h_\varepsilon(t)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t \\
=&\int_0^s \int_{\mathbb{R}^d} (\partial_t \varphi(t,x)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t + \int_s^{s+\varepsilon} \int_{\mathbb{R}^d} (\partial_t \varphi(t,x)) h_\varepsilon(t) - \varphi(t,x) \frac{1}{\varepsilon}\,\mathrm{d}\mu_t(x)\,\mathrm{d}t \\
=&A_1 + A_2 + A_3.
\end{aligned}
$$

Note that $(\partial_t \varphi(t,x)) h_\varepsilon(t)$ is bounded due to the Lipschitz continuity of φ, so that $A_2 \to 0$ for $\varepsilon \to 0$ by dominated convergence. For A_3 we apply Lebesgue's Differentiation Theorem and use narrow continuity of $\mu_\bullet$ which leads to

$$
A_3 = -\frac{1}{\varepsilon} \int_s^{s+\varepsilon} \int_{\mathbb{R}^d} \varphi(t,x)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t \to - \int_{\mathbb{R}^d} \varphi(s,x)\,\mathrm{d}\mu_s(x),
$$

so that in total we have the convergence

$$
\int_0^T \int_{\mathbb{R}^d} \partial_t \left(\varphi(t,x) h_\varepsilon(t)\right) \mathrm{d}\mu_t(x)\,\mathrm{d}t \to \int_0^s \int_{\mathbb{R}^d} (\partial_t \varphi(t,x)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t - \int_{\mathbb{R}^d} \varphi(s,x)\,\mathrm{d}\mu_s(x). \tag{3.1.11}
$$

The other terms on the RHS of (3.1.9) are actually simpler as they do not involve time derivatives. We use Dominated Convergence Theorem again which yields

$$
\begin{aligned}
&\int_0^T \int_{\mathbb{R}^d} \nabla_x \left(\varphi(t,x) h_\varepsilon(t)\right) \cdot b(t,x) + \varphi(t,x) h_\varepsilon(t)\, c(t,x)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t \\
=&\int_0^s \int_{\mathbb{R}^d} \nabla_x \left(\varphi(t,x)\right) \cdot b(t,x) + \varphi(t,x)\, c(t,x)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t \\
&+ \int_s^{s+\varepsilon} \int_{\mathbb{R}^d} h_\varepsilon(t) \nabla_x \left(\varphi(t,x)\right) \cdot b(t,x) + \varphi(t,x) h_\varepsilon(t)\, c(t,x)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t \\
\to&\int_0^s \int_{\mathbb{R}^d} \nabla_x \left(\varphi(t,x)\right) \cdot b(t,x) + \varphi(t,x)\, c(t,x)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t \qquad (\varepsilon \to 0).
\end{aligned} \tag{3.1.12}
$$

Similarly, one can show that for $\varepsilon \to 0$

$$\begin{aligned}&\int_0^T \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \varphi(t,y) h_\varepsilon(t) \,\mathrm{d}[\eta(t,x)]\,(y) \right) \mathrm{d}\mu_t(x)\,\mathrm{d}t + \int_0^T \int_{\mathbb{R}^d} \varphi(t,x) h_\varepsilon(t)\,\mathrm{d}[N(t)](x)\,\mathrm{d}t \\ \to &\int_0^s \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \varphi(t,y)\,\mathrm{d}[\eta(t,x)]\,(y) \right) \mathrm{d}\mu_t(x)\,\mathrm{d}t + \int_0^s \int_{\mathbb{R}^d} \varphi(t,x)\,\mathrm{d}[N(t)](x)\,\mathrm{d}t.\end{aligned} \tag{3.1.13}$$

Combining (3.1.10) with (3.1.11), (3.1.12) and (3.1.13) yields the claim. □

In order to construct a unique solution to (3.1.1) and to prove continuous dependency on model parameters as well as initial measures we proceed in several steps. First, in Section 3.2 we prepare basic statemens on the flow of the vector field b. As a second step, in Section 3.3 we set the state-independent influx N to zero, so that the PDE (3.1.1) is linear in μ_t. This allows us to move to the adjoint problem by duality theory and to explicitly find a unique solution with the method of characteristics and Banach Fixed Point Theorem. In Section 3.4, we apply Duhamel's principle to construct a solution to problems with non-vanishing influx N.
The general nonlinear version of (3.1.1) in which the model functions can depend on the solution μ_t is treated in Section 3.5.

3.2 The flow of a vector field

In the construction of the measure solution we will apply the flow which is generated by the vector field b. To this end, we collect here several basic properties and estimates concerning the flow so that they don't disturb the presentation in the next sections.

Definition 3.2.1. *Let b as in Assumptions 3.1.1 and let $s, \tau \in [0,T]$. Then we denote the unique solution of the ODE*

$$\partial_s X_b(s;\tau,x) = b(s, X_b(s;\tau,x)) \qquad X_b(\tau;\tau,x) = x \tag{3.2.1}$$

*by $X_b : [0,T] \times [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ and call it the **flow** of the vector field b.*

Remark 3.2.2. Basic results from ODE theory guarantee that X_b indeed exists and is unique. In particular, Picard-Lindelöf Theorem (cf. [161, II.§6.I]) gives a locally unique C^1 solution $s \mapsto X_b(s;\tau,x)$ which according to [149, Cororllary 2.6] can be extended to the whole interval $[0, T]$ as

$$\int_0^T |b(s,\cdot)|_{\mathbf{Lip}}\,\mathrm{d}s \le \int_0^T \|b(s,\cdot)\|_{BL}\,\mathrm{d}s \le T\|b\|_\infty < \infty.$$

The flow ist Lipschitz continuous in all arguments as we show in the next lemma. Since we will later restrict the domain of the flow to intervals of the form $[0, t]$ with $0 \le t \le T$, we present a slightly more general result to get sharper estimates.

Lemma 3.2.3. *Let b as in Assumptions 3.1.1 and let $t \in [0, T]$ be arbitrary. Then the following statements hold*

i) For all $s, \tau \in [0, t]$ the map $\mathbb{R}^d \ni x \mapsto X_b(s;\tau,x)$ is Lipschitz continuous, i.e.

$$|X_b(s;\tau,x_1) - X_b(s;\tau,x_2)| \le |x_1 - x_2| e^{t\|b\|_\infty}. \tag{3.2.2}$$

ii) For all $(\tau, x) \in [0, t] \times \mathbb{R}^d$ the map $[0, t] \ni s \mapsto X_b(s;\tau,x)$ is Lipschitz continuous, i.e.

$$|X_b(s_1;\tau,x) - X_b(s_2;\tau,x)| \le \|b\|_\infty |s_1 - s_2|. \tag{3.2.3}$$

iii) For all $(s, x) \in [0, t] \times \mathbb{R}^d$ the map $[0, t] \ni \tau \mapsto X_b(s;\tau,x)$ is Lipschitz continuous, i.e.

$$|X_b(s;\tau_1,x) - X_b(s;\tau_2,x)| \le \|b\|_\infty e^{t\|b\|_\infty} |\tau_1 - \tau_2|. \tag{3.2.4}$$

Proof. i) We have

$$\begin{aligned}
& X_b(s;\tau,x_1) - X_b(s;\tau,x_2) \\
=& X_b(s;\tau,x_1) - X_b(\tau;\tau,x_1) + X_b(\tau;\tau,x_2) - X_b(s;\tau,x_2) + x_1 - x_2 \\
=& \int_\tau^s \partial_\sigma [X_b(\sigma;\tau,x_1) - X_b(\sigma;\tau,x_2)]\,\mathrm{d}\sigma + x_1 - x_2 \\
=& \int_\tau^s b(\sigma, X_b(\sigma;\tau,x_1)) - b(\sigma, X_b(\sigma;\tau,x_2))\,\mathrm{d}\sigma + x_1 - x_2.
\end{aligned}$$

If $\tau \le s$, then

$$|X_b(s;\tau,x_1) - X_b(s;\tau,x_2)| \le \int_\tau^s |b(\sigma,\cdot)|_{\mathbf{Lip}} |X_b(\sigma;\tau,x_1) - X_b(\sigma;\tau,x_2)|\,\mathrm{d}\sigma + |x_1 - x_2|,$$

so that Gronwall's inequality yields

$$|X_b(s;\tau,x_1) - X_b(s;\tau,x_2)| \leq |x_1 - x_2| e^{\int_\tau^s |b(\sigma,\cdot)|_{\mathbf{Lip}}\,\mathrm{d}\sigma} \leq |x_1 - x_2| e^{t\|b\|_\infty}.$$

If $\tau > s$, we have

$$|X_b(s;\tau,x_1) - X_b(s;\tau,x_2)| \leq \int_s^\tau |b(\sigma,\cdot)|_{\mathbf{Lip}} |X_b(\sigma;\tau,x_1) - X_b(\sigma;\tau,x_2)|\,\mathrm{d}\sigma + |x_1 - x_2|,$$

and we can proceed analogously to get the same bound. Similarly, X_b is Lipschitz continuous in s since for $s_1 \leq s_2$

$$|X_b(s_2;\tau,x) - X_b(s_1;\tau,x)| = \left| \int_{s_1}^{s_2} |b(\sigma, X_b(\sigma;\tau,x))\,\mathrm{d}\sigma \right| \leq \|b\|_\infty |s_2 - s_1|,$$

which shows ii). To prove iii) let $\tau_1, \tau_2 \in [0,t]$ and without loss of generality $\tau_2 \leq \tau_1$. Then

$$\begin{aligned} &X_b(s;\tau_1,x) - X_b(s;\tau_2,x) = X_b(s;\tau_1,x) - X_b(\tau_1;\tau_1,x) + X_b(\tau_2;\tau_2,x) - X_b(s;\tau_2,x) \\ &= \int_{\tau_1}^s b(\sigma, X_b(\sigma;\tau_1,x))\,\mathrm{d}\sigma - \int_{\tau_2}^s b(\sigma, X_b(\sigma;\tau_2,x))\,\mathrm{d}\sigma \\ &= \int_{\tau_1}^s b(\sigma, X_b(\sigma;\tau_1,x)) - b(\sigma, X_b(\sigma;\tau_2,x))\,\mathrm{d}\sigma - \int_{\tau_2}^{\tau_1} b(\sigma, X_b(\sigma;\tau_2,x))\,\mathrm{d}\sigma. \end{aligned}$$

If $s \geq \tau_1$ then

$$\begin{aligned} &|X_b(s;\tau_1,x) - X_b(s;\tau_2,x)| \\ \leq &\int_s^{\tau_1} |b(\sigma,\cdot)|_{\mathbf{Lip}} |X_b(\sigma;\tau_1,x) - X_b(\sigma;\tau_2,x)|\,\mathrm{d}\sigma + (\tau_1 - \tau_2)\|b\|_\infty \\ \leq &\|b\|_\infty \int_s^{\tau_1} |X_b(\sigma;\tau_1,x) - X_b(\sigma;\tau_2,x)|\,\mathrm{d}\sigma + |\tau_1 - \tau_2| \|b\|_\infty. \end{aligned}$$

Then Gronwall's inequality yields

$$|X_b(s;\tau_1,x) - X_b(s;\tau_2,x)| \leq \|b\|_\infty |\tau_1 - \tau_2| e^{t\|b\|_\infty}.$$

The same bound can be established if $s < \tau_1$. □

Proposition 3.2.4. *Let $t \in [0,T]$ be arbitrary and let $b, \tilde{b}$ as in Assumptions 3.1.1* *with corresponding solutions* $X_b, X_{\tilde{b}}$ *to the ODE* (3.2.1)*. Then for all* $s, \tau \in [0,t]$

$$\|X_b(s;\tau,\cdot) - X_{\tilde{b}}(s;\tau,\cdot)\|_\infty \leq \|b-\tilde{b}\|_\infty \, t \, e^{2t\|\tilde{b}\|_\infty}.$$

Proof. Let $x \in \mathbb{R}^d$. With (3.2.1) and triangle inequality we see

$$\begin{aligned}
&\partial_s[X_b(s;\tau,x) - X_{\tilde{b}}(s;\tau,x)] = b(s;X_b(s;\tau,x)) - \tilde{b}(s;X_{\tilde{b}}(s;\tau,x))\\
=&b(s,X_b(s;\tau,x)) - \tilde{b}(s,X_b(s;\tau,x)) + \tilde{b}(s,X_b(s;\tau,x)) - \tilde{b}(s,X_{\tilde{b}}(s;\tau,x))\\
\leq&\|b(s,\cdot) - \tilde{b}(s,\cdot)\|_\infty + |\tilde{b}(s,\cdot)|_{\mathbf{Lip}}|X_b(s;\tau,x) - X_{\tilde{b}}(s;\tau,x)|\\
\leq&\|b-\tilde{b}\|_\infty + \|\tilde{b}\|_\infty |X_b(s;\tau,x) - X_{\tilde{b}}(s;\tau,x)|.
\end{aligned}$$

If $s \geq \tau$, then integrating from τ to s yields

$$\begin{aligned}
&X_b(s;\tau,x) - X_{\tilde{b}}(s;\tau,x) = \int_\tau^s \partial_s[X_b(\sigma;\tau,x) - X_{\tilde{b}}(\sigma;\tau,x)] \,\mathrm{d}\sigma\\
\leq&\|b-\tilde{b}\|_\infty t + \|\tilde{b}\|_\infty \int_\tau^s |X_b(\sigma;\tau,x) - X_{\tilde{b}}(\sigma,\tau,x)| \,\mathrm{d}\sigma.
\end{aligned}$$

According to a variant of the classical Gronwall's inequality [114, Theorem 3.2.1] (with $L(\sigma,x) := \|x\|$ and $M := 1$), this gives the following estimate

$$\begin{aligned}
&X_b(s;\tau,x) - X_{\tilde{b}}(s;\tau,x) \leq \|b-\tilde{b}\|_\infty t + \|\tilde{b}\|_\infty \int_\tau^s \|b-\tilde{b}\|_\infty t e^{\int_\sigma^s \|\tilde{b}\|_\infty \,\mathrm{d}\tilde{\sigma}} \,\mathrm{d}\sigma\\
\leq&\|b-\tilde{b}\|_\infty t \left[1 + \|\tilde{b}\|_\infty e^{\|\tilde{b}\|_\infty t} s\right] \leq \|b-\tilde{b}\|_\infty t e^{t\|\tilde{b}\|_\infty} \left[1 + \|\tilde{b}\|_\infty t\right] \leq \|b-\tilde{b}\|_\infty t e^{2t\|\tilde{b}\|_\infty},
\end{aligned}$$

where we used that $1 + x \leq e^x$ for all x. An application of the supremum norm yields the claim.
For $s < \tau$ we integrate from s to τ and proceed analogously to get the same bound.

$\square$

3.3 The case without state-independent influx

We are now ready to treat the linear model (3.1.1) on $[0,T] \times \mathbb{R}^d$ with the state-independent influx N set to zero, i.e.

$$\begin{cases} \partial_t \mu_t + \nabla_x \cdot (b(t,x)\mu_t) &= c(t,x)\mu_t + \int_{\mathbb{R}^d} \eta(t,x)(y) \,\mathrm{d}\mu_t(y),\\ \mu_0 &= \nu, \end{cases} \tag{3.3.1}$$

As a first step, we will establish a solution to the corresponding dual problem and then later use it to define a measure solution to the (primal) problem (3.3.1). For $t \in [0,T]$ and an arbitrary test function $\psi \in BL(\mathbb{R}^d) \cap C^1(\mathbb{R}^d)$ the dual problem corresponding to (3.3.1) is given by

$$\begin{cases} \partial_\tau \varphi_{\psi,t} + b \cdot \nabla_x \varphi_{\psi,t} + c\, \varphi_{\psi,t} + \int_{\mathbb{R}^d} \varphi_{\psi,t}(\tau,y)\, \mathrm{d}[\eta(\tau,x)]\,(y) & = 0 \quad \text{in} \quad [0,t] \times \mathbb{R}^d, \\ \varphi_{\psi,t}(t,\cdot) & = \psi \quad \text{in} \quad \mathbb{R}^d, \end{cases} \tag{3.3.2}$$

In order to solve (3.3.2), we first derive an implicit integral equation for the solution $\varphi_{\psi,t}$ which we infer from the the method of characteristics.

Proposition 3.3.1. *Let $\psi \in BL(\mathbb{R}^d)$.*

(i) There exists a bounded Lipschitz continuous solution $\tilde{\varphi}_{\psi,t}$ to (3.3.2)*.*

(ii) Let $\varphi_{\psi,t} \in BL([0,t] \times \mathbb{R}^d)$ be a solution to (3.3.2)*. Then it satisfies the following representation formula*

$$\begin{aligned} \varphi_{\psi,t}(\tau,x) =& \psi(X_b(t;\tau,x)) e^{\int_\tau^t c(r, X_b(r;\tau,x))\,\mathrm{d}r} \\ &+ \int_\tau^t \int_{\mathbb{R}^d} \varphi_{\psi,t}(s,y)\, \mathrm{d}[\eta(s, X_b(s;\tau,x))](y) e^{\int_\tau^s c(r, X_b(r;\tau,x))\,\mathrm{d}r}\, \mathrm{d}s, \end{aligned} \tag{3.3.3}$$

where X_b is the flow of b as introduced in Definition 3.2.1.

Proof. Before we treat the terminal value problem (3.3.2), we first consider the following initial value problem analysed in Ref. [31]. For $t > 0$ let

$$\begin{cases} \partial_\tau \xi(\tau,x) + A(\tau,x) \cdot \nabla_x \xi(\tau,x) - B(\tau,x)\xi(\tau,x) - C(\tau,x) & = 0 \quad \text{in } [0,t] \times \mathbb{R}^d \\ \xi(0,\cdot) & = \psi \quad \text{in} \quad \mathbb{R}^d. \end{cases}, \tag{3.3.4}$$

Assume that the model functions A, B, C satisfy

(a) $A \in C^0([0,t]; BL(\mathbb{R}^d; \mathbb{R}^d))$,

(b) A satisfies *condition M*, i.e. there exists a nonnegative function $K \in L^1([0,T])$ such that

$$\langle x - y, A(\tau,x) - A(\tau,y) \rangle \geq -K(t) \|x - y\|^2 \quad \forall (\tau,x), (\tau,y) \in [0,t] \times \mathbb{R}^d,$$

(c) $B, C \in BL([0,t] \times \mathbb{R}^d)$.

As we assumed A to be continuous [31, Corollary] implies that there exists a bounded Lipschitz function $\xi : [0,t] \times \mathbb{R}^d \to \mathbb{R}$ solving (3.3.4) λ^{d+1} almost everywhere and which is given explicitly by

$$\begin{aligned}\xi(\tau,x) =& \psi\left(X_A(0;\tau,x)\right)\ e^{\int_0^\tau\ B(r,X_A(r;\tau,x))\,\mathrm{d}r} \\ &+ \int_0^\tau\ C\left(s, X_A(s;\tau,x)\right)\ e^{\int_s^\tau B(r,X_A(r;\tau,x))\,\mathrm{d}r}\,\mathrm{d}s.\end{aligned} \tag{3.3.5}$$

To derive the representation formula for $\xi(\tau,x)$ the model functions are suitably mollified so that the method of characteristics can be applied which leads to a formula in the higher regularity case. Then a limiting procedure yields (3.3.5).
By substituting $\varphi(\tau,x) := \xi(t-\tau,x)$ we construct a solution to the terminal value problem on $[0,t] \times \mathbb{R}^d$

$$\begin{cases}\partial_\tau\varphi(\tau,x) - A(t-\tau,x)\cdot\nabla_x\varphi(\tau,x) + B(t-\tau,x)\varphi(\tau,x) + C(t-\tau,x) & = 0 \\ \varphi(t,\cdot) & = \psi\end{cases}. \tag{3.3.6}$$

As before, for any $\psi \in BL(\mathbb{R}^d)$, there exists a bounded Lipschitz solution $\varphi : [0,t] \times \mathbb{R}^d \to \mathbb{R}$ solving (3.3.6) λ^{d+1} almost everywhere. This shows (i). Now we can derive a representation formula for φ via $\varphi(\tau,x) = \xi(t-\tau,x)$. To this end, we note that

$$\int_0^{t-\tau} B(r, X_A(r;t-\tau,x))\,\mathrm{d}r = \int_\tau^t B(t-r, X_A(t-r;t-\tau,x))\,\mathrm{d}r$$

and

$$\begin{aligned}&\int_0^{t-\tau} C(s, X_A(s;t-\tau,x))e^{\int_s^{t-\tau} B(r,X_A(r;t-\tau,x))\,\mathrm{d}r}\,\mathrm{d}s \\ =& \int_0^{t-\tau} C(s, X_A(s;t-\tau,x))e^{\int_\tau^{t-s} B(t-r,X_A(t-r;t-\tau,x))\,\mathrm{d}r}\,\mathrm{d}s \\ =& \int_\tau^t C(t-s, X_A(t-s;t-\tau,x))e^{\int_\tau^s B(t-r,X_A(t-r;t-\tau,x))\,\mathrm{d}r}\,\mathrm{d}s.\end{aligned}$$

Thus, using (3.3.5) the representation formula is explicitly given by

$$\begin{aligned}\varphi(\tau,x)=\xi(t-\tau,x)=&\psi(X_A(0;t-\tau,x))e^{\int_\tau^t B(t-r,X_A(t-r;t-\tau,x))\,\mathrm{d}r}\\&+\int_\tau^t C(t-s,X_A(t-s;t-\tau,x))e^{\int_\tau^s B(t-r,X_A(t-r;t-\tau,x))\,\mathrm{d}r}\,\mathrm{d}s.\end{aligned}$$

Now we set

$$\begin{aligned}A(t-\tau,x)&:=-b(\tau,x),\\B(t-\tau,x)&:=c(\tau,x),\\C(t-\tau,x)&:=\int_{\mathbb{R}^d}\varphi(\tau,y)\,\mathrm{d}[\eta(\tau,x)](y).\end{aligned}\tag{3.3.7}$$

By Assumption 3.1.1 the model functions b,c,η provide enough regularity that assumptions (a) and (c) are satisfied so that we are left to check assumption (b), i.e. *condition M*. To this end let $(\tau,x),(\tau,y)\in[0,t]\times\mathbb{R}^d$, then using the Lipschitz continuity of b

$$\begin{aligned}&\langle x-y,A(\tau,x)-A(\tau,y)\rangle=\langle x-y,-b(t-\tau,x)+b(t-\tau,y)\rangle\\=&\sum_{i=1}^d(x_i-y_i)(b_i(t-\tau,y)-b_i(t-\tau,x))\geq-\sum_{i=1}^d|x_i-y_i||b_i(t-\tau,y)-b_i(t-\tau,x)|\\\geq&-\sum_{i=1}^d|x_i-y_i||b_i(t-\tau,\cdot)|_{\mathbf{Lip}}|x-y|\geq-\|b\|_\infty|x-y|^2.\end{aligned}$$

We are left to check how the representation formula changes with the specific choices of the model functions. In particular, we have to transform X_A accordingly. Uniqueness of the ODE (3.2.1) $s\in[0,T]$ yields that

$$X_A(\tau;s,x)=X_b(t-\tau;t-s,x),\tag{3.3.8}$$

which can be checked quickly. Plugging (3.3.7) into model (3.3.6) leads to the PDE problem (3.3.2). The the corresponding representation formula can be transformed using (3.3.8) which gives

$$\begin{aligned}\varphi_{\psi,t}(\tau,x)=&\psi(X_b(t;\tau,x))e^{\int_\tau^t c(r,X_b(r;\tau,x))\,\mathrm{d}r}\\&+\int_\tau^t\int_{\mathbb{R}^d}\varphi_{\psi,t}(s,y)\,\mathrm{d}[\eta(s,X_b(s;\tau,x))](y)e^{\int_\tau^s c(r,X_b(r;\tau,x))\,\mathrm{d}r}\,\mathrm{d}s\end{aligned}$$

as desired. $\square$

Remark 3.3.2. Since all model functions are actually defined on $[0,T]$ and hence also X_b by Remark 3.2.2, any function $\varphi_{\psi,t}$ defined by the representation formula (3.3.3) can be naturally extended to the whole interval $[0,T]$. Note however, that the extension will not necessarily be related to the dual problem (3.3.4) on $[t,T]$.

As a first step in constructing a solution to the dual problem (3.3.2), we construct a unique solution to the representation formula (3.3.3) on $[0,T]$ via Banach Fixed Point Theorem.

Proposition 3.3.3. *Let $t \in [0,T]$ and $\psi \in C^0(\mathbb{R}^d)$. Then there exists a unique solution $\varphi_{\psi,t} \in C^0([0,T]\times\mathbb{R}^d)$ to the implicit integral representation* (3.3.3).

Proof. We want to argue with Banach Fixed Point Theorem. To this end, consider the operator $L : C^0([0,T]\times\mathbb{R}^d) \to C^0([0,T]\times\mathbb{R}^d)$ given by the right-hand side of (3.3.3), i.e. for $g \in C^0([0,T]\times\mathbb{R}^d)$

$$\begin{aligned}L(g)(\tau,x) =&\psi(X_b(t;\tau,x))e^{\int_\tau^t c(r,X_b(r;\tau,x))\,\mathrm{d}r}\\ &+\int_\tau^t\int_{\mathbb{R}^d} g(s,y)\,\mathrm{d}[\eta(s,X_b(s;\tau,x))](y)e^{\int_\tau^s c(r,X_b(r;\tau,x))\,\mathrm{d}r}\,\mathrm{d}s.\end{aligned}$$

Due to the continuity of all the model functions as well as of ψ and X_b (cf. Lemma 3.2.3) it is a standard computation to check that $L(g)$ is indeed continuous.
In the following, for $\lambda > 0$ we use the Bielecki norm on $C^0([0,T]\times\mathbb{R}^d)$

$$\|g\|_\lambda = \sup_{(\tau,x)\in[0,T]\times\mathbb{R}^d} e^{-\lambda(T-\tau)}|g(\tau,x)| \tag{3.3.9}$$

A short computation reveals that $\|\cdot\|_\lambda$ is equivalent to the norm $\|\cdot\|_\infty$.

To show that L is a contraction let $g_1, g_2 \in C^0([0,T] \times \mathbb{R}^d)$.

$$\begin{aligned}
&\|L(g_1) - L(g_2)\|_{\lambda,f} \\
=& \sup_{(\tau,x)\in[0,T]\times\mathbb{R}^d} e^{-\lambda(T-\tau)} \\
&\cdot \left| \int_\tau^t \int_{\mathbb{R}^d} g_1(s,y) - g_2(s,y) \,\mathrm{d}[\eta(s, X_b(s;\tau,x))](y) e^{\int_\tau^s c(r,X_b(r;\tau,x))\,\mathrm{d}r} \,\mathrm{d}s \right| \\
\leq& e^{T\|c\|_\infty} \sup_{(\tau,x)\in[0,T]\times\mathbb{R}^d} e^{-\lambda(T-\tau)} \int_\tau^t \int_{\mathbb{R}^d} |g_1(s,y) - g_2(s,y)| \,\mathrm{d}[\eta(s, X_b(s;\tau,x))](y) \,\mathrm{d}s \\
\leq& e^{T\|c\|_\infty} \sup_{(\tau,x)\in[0,T]\times\mathbb{R}^d} e^{-\lambda(T-\tau)} \int_\tau^t \int_{\mathbb{R}^d} \|g_1 - g_2\|_\lambda e^{\lambda(T-s)} \,\mathrm{d}[\eta(s, X_b(s;\tau,x))](y) \,\mathrm{d}s \\
\leq& e^{T\|c\|_\infty} \|g_1 - g_2\|_\lambda \|\eta\|_\infty \sup_{\tau\in[0,T]} e^{\lambda\tau} \int_\tau^t e^{-\lambda s} \,\mathrm{d}s \\
=& e^{T\|c\|_\infty} \|g_1 - g_2\|_\lambda \|\eta\|_\infty \sup_{\tau\in[0,T]} e^{\lambda\tau} \frac{e^{-\lambda\tau} - e^{-\lambda t}}{\lambda} \\
=& \frac{e^{T\|c\|_\infty}\|\eta\|_\infty}{\lambda} \|g_1 - g_2\|_\lambda \sup_{\tau\in[0,T]} (1 - e^{-\lambda(t-\tau)}) \leq \frac{e^{T\|c\|_\infty}\|\eta\|_\infty}{\lambda} \|g_1 - g_2\|_\lambda .
\end{aligned}$$

So if we choose $\lambda > 2e^{T\|c\|_\infty}\|\eta\|_\infty$ then L is a contraction with respect to $\|\cdot\|_\lambda$ and hence we conclude by Banach Fixed Point Theorem. □

Now that we established a unique solution $\varphi_{\psi,t}$ to the implicit representation formula (3.3.3), we will concentrate on constructing a solution to the dual problem (3.3.2). To this end, we restrict the domain of τ to $[0,t]$ where $\varphi_{\psi,t}$ actually corresponds to the PDE, i.e. we consider $\varphi_{\psi,t} \in C^0([0,t] \times \mathbb{R}^d)$. Nevertheless, it will be quite helpful later to remember that $\varphi_{\psi,t}$ is defined for all $\tau \in [0,T]$.

Proposition 3.3.4. *Let $\psi \in BL(\mathbb{R}^d)$ and $t \in [0,T]$. Then the unique solution $\varphi_{\psi,t} \in C^0([0,t] \times \mathbb{R}^d)$ to the implicit representation formula* (3.3.3) *is bounded by*

$$\|\varphi_{\psi,t}\|_\infty \leq \|\psi\|_\infty e^{t(\|c\|_\infty + \|\eta\|_\infty)}.$$

Proof. We first bound the x-component. If $\tau \in [0,t]$, then

$$|\varphi_{\psi,t}(\tau,x)| \leq \|\psi\|_\infty e^{t\|c\|_\infty} + \int_\tau^t \int_{\mathbb{R}^d} |\varphi_{\psi,t}(s,y)| \,\mathrm{d}[\eta(s, X_b(s;\tau,x))](y) e^{(s-\tau)\|c\|_\infty} \,\mathrm{d}s,$$

so that

$$\begin{aligned}\sup_{\tilde{x}\in\mathbb{R}^d}|\varphi_{\psi,t}(\tau,\tilde{x})| \leq& \|\psi\|_\infty e^{t\|c\|_\infty} + \int_\tau^t \sup_{\tilde{x}\in\mathbb{R}^d}\int_{\mathbb{R}^d}|\varphi_{\psi,t}(s,y)|\,\mathrm{d}[\eta(s,\tilde{x})](y)e^{(s-\tau)\|c\|_\infty}\,\mathrm{d}s\\
\leq& \|\psi\|_\infty e^{t\|c\|_\infty} + \int_\tau^t \sup_{\tilde{y}\in\mathbb{R}^d}|\varphi_{\psi,t}(s,\tilde{y})| \sup_{\tilde{x}\in\mathbb{R}^d}\int_{\mathbb{R}^d}\mathrm{d}[\eta(s,\tilde{x})](y)e^{(s-\tau)\|c\|_\infty}\,\mathrm{d}s\\
\leq& \|\psi\|_\infty e^{t\|c\|_\infty} + \int_\tau^t \sup_{\tilde{y}\in\mathbb{R}^d}|\varphi_{\psi,t}(s,\tilde{y})|\|\eta\|_\infty e^{(s-\tau)\|c\|_\infty}\,\mathrm{d}s\\
\leq& \|\psi\|_\infty e^{t\|c\|_\infty} + \|\eta\|_\infty e^{-\tau\|c\|_\infty}\int_\tau^t \sup_{\tilde{y}\in\mathbb{R}^d}|\varphi_{\psi,t}(s,\tilde{y})|e^{s\|c\|_\infty}\,\mathrm{d}s.\end{aligned}$$

Note that the map $\tau \mapsto \|\psi\|_\infty e^{(t+\tau)\|c\|_\infty}$ is monotonically increasing, so that with Gronwall's inequality applied to $\tau \mapsto e^{\tau\|c\|_\infty}\sup_{\tilde{x}\in\mathbb{R}^d}|\varphi_{\psi,t}(\tau,\tilde{x})|$ we conclude

$$e^{\tau\|c\|_\infty}\sup_{\tilde{x}\in\mathbb{R}^d}|\varphi_{\psi,t}(\tau,\tilde{x})| \leq \|\psi\|_\infty e^{(t+\tau)\|c\|_\infty}e^{(t-\tau)\|\eta\|_\infty} \leq \|\psi\|_\infty e^{(t+\tau)\|c\|_\infty}e^{t\|\eta\|_\infty},$$

or equivalently

$$\sup_{\tilde{x}\in\mathbb{R}^d}|\varphi_{\psi,t}(\tau,\tilde{x})| \leq \|\psi\|_\infty e^{t(\|c\|_\infty+\|\eta\|_\infty)}.$$

□

If we consider $\varphi_{\psi,t}$ as a function on $[0,T]\times\mathbb{R}^d$ then we can get a similar bound which is independent of t.

Corollary 3.3.5. *Let $\psi \in BL(\mathbb{R}^d)$ and $t \in [0,T]$. Then $\varphi_{\psi,t} \in C^0([0,T]\times\mathbb{R}^d)$ can be bound independently of t by*

$$\|\varphi_{\psi,t}\|_\infty \leq \|\psi\|_\infty e^{T(\|c\|_\infty+\|\eta\|_\infty)}.$$

Proof. In view of Proposition 3.3.4 we have already a sharper bound $\varphi_{\psi,t}$ if $\tau \leq t$. In the case of $t < \tau \leq T$ we can get a slightly weaker but similar estimate of the form

$$\sup_{\tilde{x}\in\mathbb{R}^d}|\varphi_{\psi,t}(\tau,\tilde{x})| \leq \|\psi\|_\infty e^{T\|c\|_\infty} + \|\eta\|_\infty e^{-\tau\|c\|_\infty}\int_t^\tau \sup_{\tilde{y}\in\mathbb{R}^d}|\varphi_{\psi,t}(s,\tilde{y})|e^{s\|c\|_\infty}\,\mathrm{d}s,$$

which with Gronwall's inequality applied to $\tau \mapsto e^{\tau\|c\|_\infty} \sup_{\tilde{x}\in\mathbb{R}^d} |\varphi_{\psi,t}(\tau,\tilde{x})|$ leads to

$$e^{\tau\|c\|_\infty} \sup_{\tilde{x}\in\mathbb{R}^d} |\varphi_{\psi,t}(\tau,\tilde{x})| \leq \|\psi\|_\infty e^{(T+\tau)\|c\|_\infty} e^{(\tau-t)\|\eta\|_\infty} \leq \|\psi\|_\infty e^{(T+\tau)\|c\|_\infty} e^{T\|\eta\|_\infty}$$

or reformulated to

$$\sup_{\tilde{x}\in\mathbb{R}^d} |\varphi_{\psi,t}(\tau,\tilde{x})| \leq \|\psi\|_\infty e^{T(\|c\|_\infty+\|\eta\|_\infty)}.$$

In particular, we get a uniform bound for all $\tau \in [0,T]$ and thus the statement follows. □

Our next step is to establish Lipschitz continuity of $\varphi_{\psi,t}$ and for this we need the following regularity statements which are based on [44, Lemma 3.17], however with upgraded regularity in time of the model functions. For the convenience of the reader, we sketch the arguments.

Lemma 3.3.6. *Let $\psi \in BL(\mathbb{R}^d)$ and define*

$$G_\tau^{\psi,t} : \mathbb{R}^d \to \mathbb{R}, \qquad x \mapsto \psi(X_b(t;\tau,x)) e^{\int_\tau^t c(r,X_b(r;\tau,x))\,\mathrm{d}r}.$$

Then the following statements hold:

i) For $t \in [0,T]$ and $\tau \in [0,t]$ the map $x \mapsto G_\tau^{\psi,t}(x)$ is in $BL(\mathbb{R}^d)$ with

$$|G_\tau^{\psi,t}(x_1) - G_\tau^{\psi,t}(x_2)| \leq \|\psi\|_{BL} e^{t(\|b\|_\infty+2\|c\|_\infty)} |x_1 - x_2|. \tag{3.3.10}$$

ii) For $t \in [0,T]$ and all $x \in \mathbb{R}^d$ the map $[0,t] \ni \tau \mapsto G_\tau^{\psi,t}(x)$ is Lipschitz continuous with

$$|G_{\tau_1}^{\psi,t}(x) - G_{\tau_2}^{\psi,t}(x)| \leq \|\psi\|_{BL}(\|b\|_\infty + \|c\|_\infty) e^{t(2\|b\|_\infty+\|c\|_\infty)} |\tau_1 - \tau_2|. \tag{3.3.11}$$

iii) For all $\tau \in [0,T]$ and all $x \in \mathbb{R}^d$ the map $[0,T] \ni t \mapsto G_\tau^{\psi,t}$ is Lipschitz continuous with

$$|G_\tau^{\psi,t_1}(x) - G_\tau^{\psi,t_2}(x)| \leq \|\psi\|_{BL}(\|b\|_\infty + \|c\|_\infty) e^{T\|c\|_\infty} |t_1 - t_2|. \tag{3.3.12}$$

Proof. Before we show i), we note that the map $x \mapsto e^x$ is Lipschitz continuous on compact sets K with Lipschitz constant bounded by $e^{\sup_{x\in K}|x|}$. Hence, we compute with (3.2.2) for $x_1, x_2 \in \mathbb{R}^d$

$$\begin{aligned}
&\left| e^{\int_\tau^t c(r, X_b(r;\tau,x_1))\,\mathrm{d}r} - e^{\int_\tau^t c(r, X_b(r;\tau,x_2))\,\mathrm{d}r} \right| \\
\leq & e^{t\|c\|_\infty} \int_\tau^t |c(r, X_b(r;\tau,x_1)) - c(r, X_b(r;\tau,x_2))|\,\mathrm{d}r \\
\leq & e^{t\|c\|_\infty} \int_\tau^t |c(r,\cdot)|_{\mathbf{Lip}} |X_b(r;\tau,x_1) - X_b(r;\tau,x_2)|\,\mathrm{d}r \\
\leq & e^{t\|c\|_\infty} \|c\|_\infty\, t\, e^{t\|b\|_\infty} |x_1 - x_2| \leq e^{t(\|b\|_\infty + \|c\|_\infty)} \|c\|_\infty\, t |x_1 - x_2|.
\end{aligned} \tag{3.3.13}$$

Now we turn to the Lipschitz continuity of $G_\tau^{\psi,t}$ in x: With (3.2.2) and (3.3.13),

$$\begin{aligned}
|G_\tau^{\psi,t}(x_1) - G_\tau^{\psi,t}(x_2)| \leq & |x_1 - x_2| \left[|\psi|_{\mathbf{Lip}} e^{t\|b\|_\infty} e^{t\|c\|_\infty} + \|\psi\|_\infty e^{t(\|b\|_\infty + \|c\|_\infty)} \|c\|_\infty\, t \right] \\
\leq & |x_1 - x_2| \|\psi\|_{BL} e^{t(\|c\|_\infty + \|b\|_\infty)} \left[1 + \|c\|_\infty t \right] \\
\leq & |x_1 - x_2| \|\psi\|_{BL} e^{t(2\|c\|_\infty + \|b\|_\infty)},
\end{aligned}$$

which shows i). Next, we show ii). From (3.2.4) we conclude

$$|\psi(X_b(t;\tau_1,x)) - \psi(X_b(t;\tau_2,x))| \leq |\psi|_{\mathbf{Lip}} \|b\|_\infty e^{t\|b\|_\infty} |\tau_1 - \tau_2|, \tag{3.3.14}$$

so that similarly to i) we get for all $\tau_1, \tau_2 \in [0,t]$

$$\begin{aligned}
&\left| e^{\int_{\tau_1}^t c(r, X_b(r;\tau_1,x))\,\mathrm{d}r} - e^{\int_{\tau_2}^t c(r, X_b(r;\tau_2,x))\,\mathrm{d}r} \right| \\
\leq & e^{t\|c\|_\infty} \left| \int_{\tau_1}^t c(r, X_b(r;\tau_1,x))\,\mathrm{d}r - \int_{\tau_2}^t c(r, X_b(r;\tau_2,x))\,\mathrm{d}r \right| \\
\leq & e^{t\|c\|_\infty} \left[|\tau_2 - \tau_1| \|c\|_\infty + \int_{\tau_2}^t \|c\|_\infty |X_b(r;\tau_1,x) - X_b(r;\tau_2,x)|\,\mathrm{d}r \right] \\
\leq & e^{t\|c\|_\infty} \left[|\tau_2 - \tau_1| \|c\|_\infty + t\|c\|_\infty \|b\|_\infty e^{t\|b\|_\infty} |\tau_1 - \tau_2| \right] \\
\leq & e^{t\|c\|_\infty} |\tau_2 - \tau_1| \|c\|_\infty e^{t\|b\|_\infty} (1 + t\|b\|_\infty) \leq \|c\|_\infty e^{t\|b\|_\infty} e^{t(\|c\|_\infty + \|b\|_\infty)} |\tau_2 - \tau_1| \\
\leq & \|c\|_\infty e^{t(\|c\|_\infty + 2\|b\|_\infty)} |\tau_2 - \tau_1|.
\end{aligned} \tag{3.3.15}$$

Note that we used the estimate $1 + x \leq e^x$ in the last step. From (3.3.14) and (3.3.15) we conclude

$$\begin{aligned}
|G_{\tau_1}^{\psi,t}(x) - G_{\tau_2}^{\psi,t}(x)| \leq & \left[|\psi|_{\mathbf{Lip}} \|b\|_\infty e^{t\|b\|_\infty} e^{t\|c\|_\infty} + \|\psi\|_\infty \|c\|_\infty e^{t(\|c\|_\infty + 2\|b\|_\infty)} \right] |\tau_2 - \tau_1| \\
\leq & \|\psi\|_{BL} (\|b\|_\infty + \|c\|_\infty) e^{t(\|c\|_\infty + 2\|b\|_\infty)} |\tau_2 - \tau_1|,
\end{aligned}$$

which was to show. To prove iii) we compute for $t_1, t_2 \in [0, T]$

$$\begin{aligned} &\left| e^{\int_\tau^{t_1} c(r, X_b(r;\tau,x))\,\mathrm{d}r} - e^{\int_\tau^{t_2} c(r, X_b(r;\tau,x))\,\mathrm{d}r} \right| \\ \le & e^{T\|c\|_\infty} \left| \int_\tau^{t_1} c(r, X_b(r;\tau,x))\,\mathrm{d}r - \int_\tau^{t_2} c(r, X_b(r;\tau,x))\,\mathrm{d}r \right| \\ \le & e^{T\|c\|_\infty} |t_1 - t_2| \|c\|_\infty . \end{aligned} \tag{3.3.16}$$

Furthermore, estimate (3.2.3) for $t = T$ yields

$$|\psi(X_b(t_1;\tau,x)) - \psi(X_b(t_2;\tau,x))| \le |\psi|_{\mathbf{Lip}} \|b\|_\infty |\tau_1 - \tau_2|, \tag{3.3.17}$$

so that with (3.3.16)

$$\begin{aligned} |G_\tau^{\psi,t_1}(x) - G_\tau^{\psi,t_2}(x)| \le & |\psi|_{\mathbf{Lip}} \|b\|_\infty |t_1 - t_2| + \|\psi\|_\infty e^{T\|c\|_\infty} |t_1 - t_2| \|c\|_\infty \\ \le & \|\psi\|_{BL} (\|b\|_\infty + \|c\|_\infty) e^{T\|c\|_\infty} |t_1 - t_2|. \end{aligned}$$

□

Proposition 3.3.7. *Let $\psi \in BL(\mathbb{R}^d)$ and $t \in [0, T]$. Then $\varphi_{\psi,t} \in BL([0,t] \times \mathbb{R}^d)$, i.e. Lipschitz continuous in τ and x, with*

$$\|\varphi_{\psi,t}\|_{BL} \le C \|\psi\|_{BL} e^{Ct}.$$

Proof. We first prove Lipschitz continuity in x. Let $\tau \in [0, t]$. First, with (3.2.2) we consider the difference

$$\begin{aligned} &\left| \int_\tau^t \int_{\mathbb{R}^d} \varphi_{\psi,t}(s,y)\,\mathrm{d}[\eta(s, X_b(s;\tau,x_1)) - \eta(s, X_b(s;\tau,x_2))](y)\,\mathrm{d}s \right| \\ \le & \|\varphi_{\psi,t}\|_\infty \int_\tau^t \|\eta(s,\cdot)\|_{BL} |X_b(s;\tau,x_1) - X_b(s;\tau,x_2)|\,\mathrm{d}s \\ \le & \|\varphi_{\psi,t}\|_\infty \|\eta\|_\infty |x_1 - x_2| \int_\tau^t e^{t\|b\|_\infty}\,\mathrm{d}s \le \|\varphi_{\psi,t}\|_\infty \|\eta\|_\infty t\, e^{t\|b\|_\infty} |x_1 - x_2|. \end{aligned} \tag{3.3.18}$$

Combining (3.3.10), (3.3.18) and (3.3.13) with some triangle inequalities yields

$$\begin{aligned}
&|\varphi_{\psi,t}(\tau,x_1)-\varphi_{\psi,t}(\tau,x_2)| \\
\leq &|x_1-x_2|\big[\|\psi\|_{BL}e^{t(\|b\|_\infty+2\|c\|_\infty)}+\|\varphi_{\psi,t}\|_\infty\|\eta\|_\infty t\, e^{t(\|b\|_\infty+\|c\|_\infty)} \\
&\qquad\qquad +\|\varphi_{\psi,t}\|_\infty\|\eta\|_\infty t\,\|c\|_\infty t\, e^{t(\|b\|_\infty+\|c\|_\infty)}\big] \\
\leq &|x_1-x_2|e^{t(\|b\|_\infty+2\|c\|_\infty)}\left[\|\psi\|_{BL}+\|\varphi_{\psi,t}\|_\infty\|\eta\|_\infty\, t+\|\varphi_{\psi,t}\|_\infty\|\eta\|_\infty\|c\|_\infty\, t^2\right].
\end{aligned}$$

Using Proposition 3.3.4 this can be simplified to

$$\begin{aligned}
&|\varphi_{\psi,t}(\tau,x_1)-\varphi_{\psi,t}(\tau,x_2)| \\
\leq &|x_1-x_2|e^{t(\|b\|_\infty+2\|c\|_\infty)}\|\psi\|_{BL}\left[1+e^{t(\|c\|_\infty+\|\eta\|_\infty)}\|\eta\|_\infty\, t+e^{t(\|c\|_\infty+\|\eta\|_\infty)}\|\eta\|_\infty\|c\|_\infty\, t^2\right] \\
\leq &|x_1-x_2|e^{t(\|b\|_\infty+3\|c\|_\infty+\|\eta\|_\infty)}\|\psi\|_{BL}\left[1+\|\eta\|_\infty\, t+\|\eta\|_\infty\|c\|_\infty\, t^2\right].
\end{aligned} \tag{3.3.19}$$

Analogously, we can prove Lipschitz continuity in τ. Let $\tau_1,\tau_2\in[0,t]$. Again, we first look at the difference

$$\begin{aligned}
&\left|\int_{\tau_1}^t\int_{\mathbb{R}^d}\varphi_{\psi,t}(s,y)\,\mathrm{d}[\eta(s,X_b(s;\tau_1,x))](y)\,\mathrm{d}s-\int_{\tau_2}^t\int_{\mathbb{R}^d}\varphi_{\psi,t}(s,y)\,\mathrm{d}[\eta(s,X_b(s;\tau_2,x))](y)\,\mathrm{d}s\right| \\
\leq &\int_{\tau_1}^{\tau_2}\int_{\mathbb{R}^d}|\varphi_{\psi,t}(s,y)|\,\mathrm{d}[\eta(s,X_b(s;\tau_1,x))](y)\,\mathrm{d}s \\
&+\left|\int_{\tau_2}^t\int_{\mathbb{R}^d}\varphi_{\psi,t}(s,y)\,\mathrm{d}[\eta(s,X_b(s;\tau_1,x))-\eta(s,X_b(s;\tau_2,x))](y)\right| \\
\leq &|\tau_1-\tau_2|\|\varphi_{\psi,t}\|_\infty\|\eta\|_\infty+\|\varphi_{\psi,t}\|_\infty\int_{\tau_2}^t\|\eta(s,\cdot)\|_{BL}|X_b(s;\tau_1,x)-X_b(s;\tau_2,x)|\,\mathrm{d}s.
\end{aligned}$$

Using (3.2.4) we continue and see

$$\begin{aligned}
&\left|\int_{\tau_1}^t\int_{\mathbb{R}^d}\varphi_{\psi,t}(s,y)\,\mathrm{d}[\eta(s,X_b(s;\tau_1,x))](y)\,\mathrm{d}s-\int_{\tau_2}^t\int_{\mathbb{R}^d}\varphi_{\psi,t}(s,y)\,\mathrm{d}[\eta(s,X_b(s;\tau_2,x))](y)\,\mathrm{d}s\right| \\
\leq &|\tau_1-\tau_2|\|\varphi_{\psi,t}\|_\infty\|\eta\|_\infty+\|\varphi_{\psi,t}\|_\infty\|\eta\|_\infty t\|b\|_\infty e^{t\|b\|_\infty}|\tau_1-\tau_2| \\
= &|\tau_1-\tau_2|\|\varphi_{\psi,t}\|_\infty\|\eta\|_\infty\left(1+t\,\|b\|_\infty e^{t\|b\|_\infty}\right) \\
\leq &|\tau_1-\tau_2|\|\varphi_{\psi,t}\|_\infty\|\eta\|_\infty e^{2t\|b\|_\infty}.
\end{aligned} \tag{3.3.20}$$

Now we can show Lipschitz continuity of $\varphi_{\psi,t}$ in τ with the estimates (3.3.11), (3.3.20) and (3.3.16)

$$
\begin{aligned}
&|\varphi_{\psi,t}(\tau_1,x)-\varphi_{\psi,t}(\tau_2,x)| \\
\leq &|\tau_1-\tau_2|\left[\|\psi\|_{BL}(\|b\|_\infty+\|c\|_\infty)e^{t(2\|b\|_\infty+\|c\|_\infty)}+\|\varphi_{\psi,t}\|_\infty\|\eta\|_\infty e^{2t\|b\|_\infty}e^{t\|c\|_\infty}\right] \\
&+|\tau_1-\tau_2|\left[\|\varphi_{\psi,t}\|_\infty\|\eta\|_\infty t\,\|c\|_\infty e^{t(2\|b\|_\infty+\|c\|_\infty)}\right] \\
\leq &|\tau_1-\tau_2|e^{t(2\|b\|_\infty+\|c\|_\infty)}\left[\|\psi\|_{BL}(\|b\|_\infty+\|c\|_\infty)+\|\varphi_{\psi,t}\|_\infty\|\eta\|_\infty+\|\varphi_{\psi,t}\|_\infty\|\eta\|_\infty t\,\|c\|_\infty\right].
\end{aligned}
$$

Again with Proposition 3.3.4 this can be simplified to

$$
\begin{aligned}
&|\varphi_{\psi,t}(\tau_1,x)-\varphi_{\psi,t}(\tau_2,x)| \\
\leq &|\tau_1-\tau_2|e^{t(2\|b\|_\infty+\|c\|_\infty)}\Bigg[\|\psi\|_{BL}(\|b\|_\infty+\|c\|_\infty)+\|\psi\|_\infty e^{t(\|c\|_\infty+\|\eta\|_\infty)}\|\eta\|_\infty \\
&\qquad +\|\psi\|_\infty e^{t(\|c\|_\infty+\|\eta\|_\infty)}\|\eta\|_\infty t\,\|c\|_\infty\Bigg] \\
\leq &|\tau_1-\tau_2|e^{t(2\|b\|_\infty+2\|c\|_\infty+\|\eta\|_\infty)}\|\psi\|_{BL}\Bigg[\|b\|_\infty+\|c\|_\infty+\|\eta\|_\infty+t\,\|\eta\|_\infty\|c\|_\infty\Bigg].
\end{aligned}
\tag{3.3.21}
$$

The boundedness follows directly from Proposition 3.3.4. □

Remark 3.3.8. With similar computations as in the proof of Proposition 3.3.7 we could even show that $\varphi_{\psi,t}\in BL([0,T]\times\mathbb{R}^d)$ if we additionally consider $\tau>t$ and replace all factors t by T. In particular, in view of the (adjusted) estimates (3.3.19), (3.3.21) and Corollary 3.3.5 the BL norm of $\varphi_{\psi,t}$ is bounded independently of t by

$$\|\varphi_{\psi,t}\|_{BL}\leq C\|\psi\|_{BL}e^{CT}$$

for some constant $C>0$ depending on T and norms of the model functions. However, for the construction of the solution to the nonlinear model in Subsection 3.5 we will need the sharper estimates from Proposition 3.3.7.

Lemma 3.3.9. *Let $\psi\in BL(\mathbb{R}^d)$. Then the map $[0,T]\ni t\mapsto\varphi_{\psi,t}$ is Lipschitz continuous, i.e. for all $(\tau,x)\in[0,T]\times\mathbb{R}^d$*

$$|\varphi_{\psi,t_1}(\tau,x)-\varphi_{\psi,t_2}(\tau,x)|\leq C\|\psi\|_{BL}|t_1-t_2|.$$

Proof. Let $t_1,t_2\in[0,T]$, without loss of generality let $t_1\geq t_2$. Then for $\tau\in[0,T]$

arbitrary

$$
\begin{aligned}
&\sup_{x\in\mathbb{R}^d} |\varphi_{\psi,t_1}(\tau,x) - \varphi_{\psi,t_2}(\tau,x)| \\
\le &\sup_{x\in\mathbb{R}^d} \left|\psi(X_b(t_1;\tau,x))e^{\int_\tau^{t_1} c(r,X_b(r;\tau,x))\,\mathrm{d}r} - \psi(X_b(t_2;\tau,x))e^{\int_\tau^{t_2} c(r,X_b(r;\tau,x))\,\mathrm{d}r}\right| \\
&+ \sup_{x\in\mathbb{R}^d} \left|\int_{t_2}^{t_1}\int_{\mathbb{R}^d} \varphi_{\psi,t_1}(s,y)\,\mathrm{d}[\eta(s,X_b(s;\tau,x))](y)e^{\int_\tau^{s} c(r,X_b(r;\tau,x))\,\mathrm{d}r}\,\mathrm{d}s\right| \\
&+ \sup_{x\in\mathbb{R}^d} \left|\int_{\tau}^{t_2}\int_{\mathbb{R}^d} \varphi_{\psi,t_1}(s,y) - \varphi_{\psi,t_2}(s,y)\,\mathrm{d}[\eta(s,X_b(s;\tau,x))](y)e^{\int_\tau^{s} c(r,X_b(r;\tau,x))\,\mathrm{d}r}\,\mathrm{d}s\right| \\
= &A + B + C.
\end{aligned}
$$

The term A can be estimated by (3.3.12)

$$
|A| \le \|\psi\|_{BL}(\|b\|_\infty + \|c\|_\infty)e^{T\|c\|_\infty}|t_1 - t_2| \le C\|\psi\|_{BL}e^{TC}|t_1 - t_2|. \tag{3.3.22}
$$

With Remark 3.3.8 we can bound the term B by

$$
|B| \le \|\varphi_{\psi,t_1}\|_{BL}\|\eta\|_\infty e^{T\|c\|_\infty}|t_1 - t_2| \le C\|\psi\|_{BL}e^{TC}|t_1 - t_2|. \tag{3.3.23}
$$

For the term C we assume $\tau \le t_2$ but notice that a comparable estimate can be established if $\tau > t_2$ by switching the integral boundaries

$$
\begin{aligned}
C \le &e^{\|c\|_\infty T} \sup_{x\in\mathbb{R}^d}\int_\tau^{t_2}\int_{\mathbb{R}^d} |\varphi_{\psi,t_1}(s,y) - \varphi_{\psi,t_2}(s,y)|\,\mathrm{d}[\eta(s,X_b(s;\tau,x))](y)\,\mathrm{d}s \\
\le &e^{\|c\|_\infty T}\|\eta\|_\infty \int_\tau^{t_2} \sup_{y\in\mathbb{R}^d} |\varphi_{\psi,t_1}(s,y) - \varphi_{\psi,t_2}(s,y)|\,\mathrm{d}s.
\end{aligned} \tag{3.3.24}
$$

So combining (3.3.22), (3.3.23) and (3.3.24) gives

$$
\begin{aligned}
&\sup_{x\in\mathbb{R}^d} |\varphi_{\psi,t_1}(\tau,x) - \varphi_{\psi,t_2}(\tau,x)| \\
\le &C|t_1 - t_2|\|\psi\|_{BL} + e^{\|c\|_\infty T}\|\eta\|_\infty \int_\tau^{t_2} \sup_{y\in\mathbb{R}^d} |\varphi_{\psi,t_1}(s,y) - \varphi_{\psi,t_2}(s,y)|\,\mathrm{d}s,
\end{aligned}
$$

which together with Gronwall's inequality leads to

$$
\begin{aligned}
&\sup_{x\in\mathbb{R}^d} |\varphi_{\psi,t_1}(\tau,x) - \varphi_{\psi,t_2}(\tau,x)| \le C|t_1 - t_2|\|\psi\|_{BL}\exp\left(e^{\|c\|_\infty T}\|\eta\|_\infty|t_2 - \tau|\right) \\
\le &C|t_1 - t_2|\|\psi\|_{BL}\exp\left(e^{\|c\|_\infty T}\|\eta\|_\infty T\right) = C\|\psi\|_{BL}|t_1 - t_2|.
\end{aligned}
$$

As the right hand-side is independent of τ the claim follows. $\square$

We conclude this paragraph with an observation.

Corollary 3.3.10. *Let $\psi \in BL(\mathbb{R}^d)$. Then the solution $\varphi_{\psi,t} \in BL([0,t] \times \mathbb{R}^d)$ to the implicit representation formula* (3.3.3) *solves the dual problem* (3.3.2) λ^{d+1} *almost everywhere.*

Proof. According to Proposition 3.3.1 (ii) any solution $\tilde{\varphi}_{\psi,t} \in BL([0,t] \times \mathbb{R}^d)$ to the dual problem (3.3.2) will also satisfy the implicit representation formula (3.3.3). However, (3.3.3) has a unique solution $\varphi_{\psi,t} \in BL([0,t] \times \mathbb{R}^d)$ (see Propositions 3.3.3 and 3.3.7). In particular, we conclude that the dual problem (3.3.2) has at most one bounded Lipschitz continuous solution, i.e. uniqueness of the problem. As existence has already been established in Proposition 3.3.1 (i), the claim follows. $\square$

Remark 3.3.11. The fact that $\varphi_{\psi,t}$ only solves (3.3.2) λ^{d+1} almost everywhere is not problematic for the construction of our measure solution. Actually, we are only interested in solving the dual problem at the boundary $\tau = t$, so that we assume $\psi \in BL(\mathbb{R}^d) \cap C^1(\mathbb{R}^d)$ and we thus have sufficient regularity where we need it.

Constructing a measure solution: Now we want to use the solution to the dual problem (3.3.2) to construct a measure solution to the primal problem (3.1.9) with $N = 0$. To this end, we plug $\varphi_{\psi,t}$ into (3.1.9) with $N = 0$ and time t which yields for the left-hand side

$$\int_{\mathbb{R}^d} \varphi_{\psi,t}(t,x)\,\mathrm{d}\mu_t(x) - \int_{\mathbb{R}^d} \varphi_{\psi,t}(0,x)\,\mathrm{d}\mu_0(x) = \int_{\mathbb{R}^d} \psi(x)\,\mathrm{d}\mu_t(x) - \int_{\mathbb{R}^d} \varphi_{\psi,t}(0,x)\,\mathrm{d}\mu_0(x)$$

and for the right-hand side

$$\begin{aligned}\int_0^t \int_{\mathbb{R}^d} &\left(\partial_\tau \varphi_{\psi,t}(\tau,x) + \nabla_x \varphi_{\psi,t}(\tau,x) \cdot b(\tau,x) + \varphi_{\psi,t}(\tau,x)\,c(\tau,x)\right) \mathrm{d}\mu_\tau(x)\,\mathrm{d}\tau \\ &+ \int_0^t \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \varphi_{\psi,t}(\tau,y)\,\mathrm{d}[\eta(\tau,x)]\,(y)\,\mathrm{d}\mu_\tau(x)\,\mathrm{d}\tau.\end{aligned} \tag{3.3.25}$$

As $\varphi_{\psi,t}$ solves the dual problem, the right-hand side (3.3.25) vanishes for almost every $x \in \mathbb{R}^d$, so we define for all $t \in [0,T]$

$$\int_{\mathbb{R}^d} \psi(x)\,\mathrm{d}\mu_t(x) = \int_{\mathbb{R}^d} \varphi_{\psi,t}(0,x)\,\mathrm{d}\mu_0(x). \tag{3.3.26}$$

Lemma 3.3.12. *Let μ_t be defined by* (3.3.26). *Then for all $\psi \in BL(\mathbb{R}^d)$ and all $0 \leq s \leq t \leq T$*

$$\int_{\mathbb{R}^d} \psi(x)\,\mathrm{d}\mu_t(x) = \int_{\mathbb{R}^d} \varphi_{\psi,t}(s,x)\,\mathrm{d}\mu_s(x). \tag{3.3.27}$$

Proof. Let $s \in [0,t]$ and let $\xi(\tau,x)$ be the solution to the dual problem (3.3.2) with final time s and boundary value $\xi(s,x) = \varphi_{\psi,t}(s,x)$. We directly conclude $\xi(\tau,x) = \varphi_{\psi,t}(\tau,x)$ for all $\tau \in [0,s]$ by uniqueness of the dual problem (see Corollary 3.3.10). Applying (3.3.26) twice yields

$$\int_{\mathbb{R}^d} \psi(x)\,\mathrm{d}\mu_t(x) = \int_{\mathbb{R}^d} \varphi_{\psi,t}(0,x)\,\mathrm{d}\mu_0(x) = \int_{\mathbb{R}^d} \xi(0,x)\,\mathrm{d}\mu_0(x) = \int_{\mathbb{R}^d} \varphi_{\psi,t}(s,x)\,\mathrm{d}\mu_s(x).$$

$\square$

Lemma 3.3.13. *Let μ_t be defined by* (3.3.26). *Then $\mu_t \in \mathcal{M}^+(\mathbb{R}^d)$ for all $t \in [0,T]$.*

Proof. In view of (3.3.26) we can prove nonnegativity of μ_t if we can show that $\psi \geq 0$ implies $\varphi_{\psi,t} \geq 0$. So using the representation formula (3.3.3) we see for arbitrary $\tau \in [0,t]$

$$\varphi_{\psi,t}(\tau,x) \geq \int_\tau^t \int_{\mathbb{R}^d} \varphi_{\psi,t}(s,y)\,\mathrm{d}[\eta(s,X_b(s;\tau,x))](y) e^{\int_\tau^s c(r,X_b(r;\tau,x))\,\mathrm{d}r}\,\mathrm{d}s. \tag{3.3.28}$$

Now we consider the convex and nonincreasing map

$$|\cdot|_- : \mathbb{R} \to \mathbb{R}_+, \qquad x \mapsto \max\{-x, 0\}$$

and apply it to both sides of (3.3.28) with Jensen's inequality

$$\begin{aligned} |\varphi_{\psi,t}(\tau,x)|_- &\leq \left| \int_\tau^t \int_{\mathbb{R}^d} \varphi_{\psi,t}(s,y)\,\mathrm{d}[\eta(s,X_b(s;\tau,x))](y) e^{\int_\tau^s c(r,X_b(r;\tau,x))\,\mathrm{d}r}\,\mathrm{d}s \right|_- \\ &\leq \int_\tau^t \int_{\mathbb{R}^d} |\varphi_{\psi,t}(s,y)|_-\,\mathrm{d}[\eta(s,X_b(s;\tau,x))](y) e^{\int_\tau^s c(r,X_b(r;\tau,x))\,\mathrm{d}r}\,\mathrm{d}s. \end{aligned} \tag{3.3.29}$$

Note that $|\lambda x|_- = \lambda |x|_-$ for $\lambda \geq 0$. Taking the supremum over x on both sides yields

$$\begin{aligned}\sup_{\tilde{x}\in\mathbb{R}^d}|\varphi_{\psi,t}(\tau,\tilde{x})|_- \leq& \sup_{\tilde{x}\in\mathbb{R}^d}\int_\tau^t\int_{\mathbb{R}^d}|\varphi_{\psi,t}(s,y)|_-\,\mathrm{d}[\eta(s,X_b(s;\tau,\tilde{x}))](y)e^{\int_\tau^s c(r,X_b(r;\tau,\tilde{x}))\,\mathrm{d}r}\,\mathrm{d}s\\ \leq& \int_\tau^t\sup_{\tilde{y}\in\mathbb{R}^d}|\varphi_{\psi,t}(s,\tilde{y})|_-\sup_{\tilde{x}\in\mathbb{R}^d}\|\eta(s,\tilde{x})\|_{BL^*}e^{\|c\|_\infty(s-\tau)}\,\mathrm{d}s\\ \leq&\|\eta\|_\infty\int_\tau^t\sup_{\tilde{y}\in\mathbb{R}^d}|\varphi_{\psi,t}(s,\tilde{y})|_-\,e^{\|c\|_\infty(s-\tau)}\,\mathrm{d}s.\end{aligned}$$

So with Gronwall's inequality we conclude $\sup_{\tilde{x}\in\mathbb{R}^d}|\varphi_{\psi,t}(\tau,\tilde{x})|_-=0$ which implies $\varphi_{\psi,t}(\tau,x)\geq 0$ for all $x\in\mathbb{R}^d$. This was to show. □

Proposition 3.3.14. *Let μ_t be defined by* (3.3.26)*. Then the following statements hold.*

i) μ_t is bounded in total variation norm, i.e.

$$\|\mu_t\|_{TV}\leq e^{t(\|c\|_\infty+\|\eta\|_\infty)}\|\mu_0\|_{TV}.$$

ii) The map $t\mapsto\mu_t$ is narrowly continuous. Furthermore, the map $t\mapsto\mu_t$ is Lipschitz continuous.

Proof. To show i) we apply the variational characterisation of the total variation norm (see e.g. [44, F.24]), (3.3.26) and Proposition 3.3.4 which yields

$$\begin{aligned}\|\mu_t\|_{TV}=&\sup_{\substack{\psi\in C_c^0(\mathbb{R}^d)\\ \|\psi\|_\infty\leq 1}}\int_{\mathbb{R}^d}\psi(x)\,\mathrm{d}\mu_t(x)\overset{(*)}{=}\sup_{\substack{\psi\in C_c^1(\mathbb{R}^d)\\ \|\psi\|_\infty\leq 1}}\int_{\mathbb{R}^d}\psi(x)\,\mathrm{d}\mu_t(x)\\ =&\sup_{\substack{\psi\in C_c^1(\mathbb{R}^d)\\ \|\psi\|_\infty\leq 1}}\int_{\mathbb{R}^d}\varphi_{\psi,t}(0,x)\,\mathrm{d}\mu_0(x)\leq\sup_{\substack{\psi\in C_c^1(\mathbb{R}^d)\\ \|\psi\|_\infty\leq 1}}\|\psi\|_\infty e^{t(\|c\|_\infty+\|\eta\|_\infty)}\|\mu_0\|_{TV}\\ =&e^{t(\|c\|_\infty+\|\eta\|_\infty)}\|\mu_0\|_{TV}.\end{aligned}$$

In $(*)$ we used that $C_c^1(\mathbb{R}^d)$ is dense in $C_0(\mathbb{R}^d)$ by Stone-Weierstrass Theorem.
Before we prove ii), we show that for all $\psi\in BL(\mathbb{R}^d)$ the map $t\mapsto\int_{\mathbb{R}^d}\psi\,\mathrm{d}\mu_t$ is

Lipschitz continuous. Indeed, let $s_1 \geq s_2$, then with Lemmata 3.3.12 and 3.3.9

$$\begin{aligned}
&\left|\int_{\mathbb{R}^d} \psi(x)\,\mathrm{d}[\mu_{s_1} - \mu_{s_2}](x)\right| = \left|\int_{\mathbb{R}^d} \varphi_{\psi,s_1}(s_2,x) - \varphi_{\psi,s_2}(s_2,x)\,\mathrm{d}\mu_{s_2}(x)\right| \\
\leq &\int_{\mathbb{R}^d} |\varphi_{\psi,s_1}(s_2,x) - \varphi_{\psi,s_2}(s_2,x)|\,\mathrm{d}\mu_{s_2}(x) \leq C|s_1 - s_2|\|\psi\|_{BL}\|\mu_{s_2}\|_{TV} \\
\leq &C|s_1 - s_2|\|\psi\|_{BL}\|\mu_0\|_{TV}.
\end{aligned} \tag{3.3.30}$$

Note that we used i) in the last step. The Lipschitz continuity of the map $t \mapsto \int_{\mathbb{R}^d} \psi\,\mathrm{d}\mu_t(x)$ directly implies that for all $t \in [0,T]$ and all $\psi \in BL(\mathbb{R}^d)$

$$\lim_{s\to t} \int_{\mathbb{R}^d} \psi(x)\,\mathrm{d}\mu_s(x) = \int_{\mathbb{R}^d} \psi(x)\,\mathrm{d}\mu_t(x),$$

which is equivalent to narrow continuity according to Portemanteau Theorem (see e.g. [44, G.14]). The Lipschitz continuity follows from (3.3.30) as well, i.e.

$$\rho_F(\mu_{s_1}, \mu_{s_2}) = \sup_{\|\psi\|_{BL}\leq 1} \int_{\mathbb{R}^d} \psi(x)\,\mathrm{d}[\mu_{s_1} - \mu_{s_2}](x) \leq C|s_1 - s_2|\|\mu_0\|_{TV}.$$

□

Our next goal is to show that the family of measures $\{\mu_t\}_{t\in[0,T]}$ actually satisfies the weak formulation. However, in the proof we will need Fréchet differentiability of $\varphi_{\psi,t}(\tau,x)$ in t. As this regularity proof is quite involved, the result is presented separately. In the following we adjust the proof of [138, 3.9] to our purposes. As a start we simplify the notation and introduce

$$\begin{aligned}
p(t,\tau,x) =&\psi(X_b(t;\tau,x))e^{\int_\tau^t c(r,X_b(r;\tau,x))\,\mathrm{d}r}, \quad g(s,\tau,x) = e^{\int_\tau^s c(r,X_b(r;\tau,x))\,\mathrm{d}r}, \\
f(t,\tau,x) =&\varphi_{\psi,t}(\tau,x),
\end{aligned} \tag{3.3.31}$$

so that we can rewrite the representation formula (3.3.3) to

$$f(t,\tau,x) = p(t,\tau,x) + \int_\tau^t \int_{\mathbb{R}^d} f(t,s,y)\,\mathrm{d}[\eta(s,X_b(s;\tau,x))](y) g(s,\tau,x)\,\mathrm{d}s. \tag{3.3.32}$$

To derive the Fréchet differentiability, we will now only work with the identity (3.3.32). As a start, we will need a generalisation of Proposition 3.3.3.

Lemma 3.3.15. *Let g as defined in* (3.3.31) *and let $q \in C_b^0([0,T]\times\mathbb{R}^d)$. For $t_0 \in [0,T]$ fixed there there exists a unique continuous map $f^* \in C_b^0([0,T]\times\mathbb{R}^d)$*

which solves the implicit equation

$$f(\tau,x)=-q(\tau,x)+\int_\tau^{t_0}\int_{\mathbb{R}^d} f(s,y)\,\mathrm{d}[\eta(s,X_b(s;\tau,x))](y)g(s,\tau,x)\,\mathrm{d}s. \tag{3.3.33}$$

Proof. We show existence and uniqueness with Banach Fixed Point Theorem and thus for $\lambda>0$ employ the Bielecki norm

$$\|f\|_\lambda=\sup_{\substack{\tau\in[0,T]\\ z\in\mathbb{R}^d}} e^{-\lambda(T-\tau)}|f(\tau,z)|.$$

Now let $L: C_b^0([0,T]\times\mathbb{R}^d)\to C_b^0([0,T]\times\mathbb{R}^d)$ be defined by

$$Lf(\tau,x)=-q(\tau,x)+\int_\tau^{t_0}\int_{\mathbb{R}^d} f(s,y)\,\mathrm{d}[\eta(s,X_b(s;\tau,x))](y)g(s,\tau,x)\,\mathrm{d}s,$$

i.e. the RHS of (3.3.33). We have

$$\begin{aligned}
&\|Lf_1-Lf_2\|_\lambda\\
=&\sup_{(\tau,z)\in[0,T]\times\mathbb{R}^d} e^{-\lambda(T-\tau)}\left|\int_\tau^{t_0}\int_{\mathbb{R}^d} f_1(s,y)-f_2(s,y)\,\mathrm{d}[\eta(s,X_b(s;\tau,z))](y)g(s,\tau,z)\,\mathrm{d}s\right|\\
\le&\|g\|_\infty \sup_{(\tau,z)\in[0,T]\times\mathbb{R}^d} e^{-\lambda(T-\tau)}\int_\tau^{t_0}\int_{\mathbb{R}^d}\|f_1-f_2\|_\lambda e^{\lambda(T-s)}\,\mathrm{d}[\eta(s,X_b(s;\tau,z))](y)\,\mathrm{d}s\\
\le&\|g\|_\infty\|\eta\|_\infty\|f_1-f_2\|_\lambda\sup_{\tau\in[0,T]} e^{\lambda\tau}\int_\tau^{t_0} e^{-\lambda s}\,\mathrm{d}s=\|g\|_\infty\|\eta\|_\infty\|f_1-f_2\|_\lambda\sup_{\tau\in[0,T]}\frac{1-e^{-\lambda(t_0-\tau)}}{\lambda}\\
\le&\frac{\|g\|_\infty\|\eta\|_\infty}{\lambda}\|f_1-f_2\|_\lambda.
\end{aligned}$$

So choosing $\lambda>2\|g\|_\infty\|\eta\|_\infty$ yields a strict contraction and (3.3.33) has a unique fixed point f^* in $C_b^0([0,T]\times\mathbb{R}^d)$. □

Now, we can focus on the Fréchet differentiability of the map $t\mapsto\varphi_{\psi,t}$.

Proposition 3.3.16. *Let f,p,g as in* (3.3.31). *Then the function $[0,T]\ni t\mapsto f(t,\cdot,\cdot)$ is Fréchet differentiable in $C_b^0([0,T]\times\mathbb{R}^d)$, i.e. for all $t_0\in[0,T]$ there exists $\partial_t f(t,\cdot,\cdot)\mid_{t=t_0}\in C_b^0([0,T]\times\mathbb{R}^d)$ such that*

$$\left\|\frac{f(t_0+\Delta t,\cdot,\cdot)-f(t_0,\cdot,\cdot)}{\Delta t}-\partial_t f(t,\cdot,\cdot)\mid_{t=t_0}\right\|_\infty\to 0 \text{ as } \Delta t\to 0.$$

Proof. Our goal is to apply the Implicit Function Theorem in Banach spaces and thus we rewrite (3.3.32) to an equation which has to be solved in $C_b^0([0,T]\times\mathbb{R}^d)$

$$\begin{aligned}0 =& -\phi(\tau,x) + p(t,\tau,x) + \int_\tau^t \int_{\mathbb{R}^d} \phi(s,y)\,\mathrm{d}[\eta(s,X_b(s;\tau,x)](y) g(s,\tau,x)\,\mathrm{d}s\\ =:& \, F(t,\phi)(\tau,x),\end{aligned}$$

where $F : [0,T]\times C_b^0([0,T]\times\mathbb{R}^d) \to C_b^0([0,T]\times\mathbb{R}^d)$ and we denoted $\phi(\tau,x) := f(t,\tau,x)$. As continuity is clear, we refer to the proof of Proposition 3.3.4 to see that $F\phi$ will be bounded for all $\phi \in C_b^0$ and thus F indeed maps into the correct function space.
As a first step of the proof, we show that F is Fréchet differentiable. Let $(\tau,x) \in [0,T]\times\mathbb{R}^d$

$$\begin{aligned}&[F(t+\Delta t,\phi+\Delta\phi) - F(t,\phi)](\tau,x)\\ =& -\phi(\tau,x) - \Delta\phi(\tau,x) + p(t+\Delta t,\tau,x)\\ &+ \int_\tau^{t+\Delta t}\int_{\mathbb{R}^d} \phi(s,y) + \Delta\phi(s,y)\,\mathrm{d}[\eta(s,X_b(s;\tau,x)](y)g(s,\tau,x)\,\mathrm{d}s\\ &+ \phi(\tau,x) - p(t,\tau,x) - \int_\tau^t\int_{\mathbb{R}^d}\phi(s,y)\,\mathrm{d}[\eta(s,X_b(s;\tau,x)](y)g(s,\tau,x)\,\mathrm{d}s\\ =& -\Delta\phi(\tau,x) + \frac{p(t+\Delta t,\tau,x)-p(t,\tau,x)}{\Delta t}\Delta t\\ &+ \int_t^{t+\Delta t}\int_{\mathbb{R}^d}\phi(s,y)\,\mathrm{d}[\eta(s,X_b(s;\tau,x)](y)g(s,\tau,x)\,\mathrm{d}s\\ &+ \int_\tau^{t+\Delta t}\int_{\mathbb{R}^d}\Delta\phi(s,y)\,\mathrm{d}[\eta(s,X_b(s;\tau,x)](y)g(s,\tau,x)\,\mathrm{d}s\\ =:& \, A+B+C.\end{aligned}$$

We consider the terms A, B and C separately. Before we consider term A, we note that p is differentiable in t since $\psi \in C^1$ and all the others functions involved also have sufficient regularity. Hence, we see

$$\begin{aligned}A =& -\Delta\phi(\tau,x) + p_t(t,\tau,x)\Delta t + \left[\frac{p(t+\Delta t,\tau,x)-p(t,\tau,x)}{\Delta t} - p_t(t,\tau,x)\right]\Delta t\\ =& -\Delta\phi(\tau,x) + p_t(t,\tau,x)\Delta t + o(\Delta t).\end{aligned} \tag{3.3.34}$$

Next, we treat term B.

$$\begin{aligned}B=&\int_t^{t+\Delta t}\underbrace{\int_{\mathbb{R}^d}\phi(s,y)\,\mathrm{d}[\eta(s,X_b(s;\tau,x))](y)g(s,\tau,x)}_{=:a(s)}\,\mathrm{d}s\\=&a(t)\Delta t+\int_t^{t+\Delta t}a(s)-a(t)\,\mathrm{d}s.\end{aligned}$$

We note that $s\mapsto a(s)$ is continuous and thus bounded on $[0,T]$. Hence,

$$\int_t^{t+\Delta t}a(s)-a(t)\,\mathrm{d}s\leq 2\|a\|_\infty\int_t^{t+\Delta t}\mathrm{d}s=2\|a\|_\infty\Delta t\in o(\Delta t).$$

In particular we get

$$B=a(t)\Delta t+o(\Delta t). \tag{3.3.35}$$

Term C can be rewritten as follows

$$\begin{aligned}C=&\int_\tau^t\int_{\mathbb{R}^d}\Delta\phi(s,y)\,\mathrm{d}[\eta(s,X_b(s;\tau,x))](y)g(s,\tau,x)\,\mathrm{d}s\\&+\int_t^{t+\Delta t}\int_{\mathbb{R}^d}\Delta\phi(s,y)\,\mathrm{d}[\eta(s,X_b(s;\tau,x))](y)g(s,\tau,x)\,\mathrm{d}s\end{aligned}$$

We observe that

$$\begin{aligned}&\int_t^{t+\Delta t}\int_{\mathbb{R}^d}\Delta\phi(s,y)\,\mathrm{d}[\eta(s,X_b(s;\tau,x))](y)g(s,\tau,x)\,\mathrm{d}s\\\leq&\|g\|_\infty\|\Delta\phi\|_\infty\|\eta\|_\infty\int_t^{t+\Delta t}\mathrm{d}s=\|g\|_\infty\|\Delta\phi\|_\infty\|\eta\|_\infty\Delta t\in o(\Delta t,\Delta\phi).\end{aligned}$$

Consequently,

$$C=\int_\tau^t\int_{\mathbb{R}^d}\Delta\phi(s,y)\,\mathrm{d}[\eta(s,X_b(s;\tau,x))](y)g(s,\tau,x)\,\mathrm{d}+o(\Delta t,\Delta\phi). \tag{3.3.36}$$

Combining the estimates (3.3.34), (3.3.35) and (3.3.36) yields

$$F(t+\Delta t,\phi+\Delta\phi)-F(t,\phi)=DF(t,\phi)[\Delta t,\Delta\phi]+o(\Delta t,\Delta\phi),$$

where $DF(t,\phi):[0,T]\times C_b^0([0,T]\times\mathbb{R}^d)\to C_b^0([0,T]\times\mathbb{R}^d)$ is a bounded linear operator defined by

$$\begin{aligned}
&DF(t,\phi)[\Delta t,\Delta\phi](\tau,x)\\
=&-\Delta\phi(\tau,x)+p_t(t,\tau,x)\Delta t+\Delta t\,a(t)+\int_\tau^t\int_{\mathbb{R}^d}\Delta\phi(s,y)\,\mathrm{d}[\eta(s,X_b(s;\tau,x)](y)g(s,\tau,x)\,\mathrm{d}s\\
=&-\Delta\phi(\tau,x)+p_t(t,\tau,x)\Delta t+\Delta t\int_{\mathbb{R}^d}\phi(t,y)\,\mathrm{d}[\eta(t,X_b(t;\tau,x)](y)g(t,\tau,x)\\
&+\int_\tau^t\int_{\mathbb{R}^d}\Delta\phi(s,y)\,\mathrm{d}[\eta(s,X_b(s;\tau,x)](y)g(s,\tau,x)\,\mathrm{d}s.
\end{aligned}$$

We remark that the continuity of the derivative $\nabla_x\psi$ secretly plays a significant role in the form of p_t for ensuring that DF maps into the correct function space.

In the second step, we want to conclude Fréchet differentiability of $\varphi_{\psi,t}$. To this end, we fix some $t_0\in[0,T]$. Our goal is to use the Implicit Function Theorem in Banach spaces [37, 5.1.29] applied to the the map $(t,\phi)\mapsto F(t,\phi)$ to express ϕ as a function of t in a small neighbourhood of t_0. We first observe that in view of Proposition 3.3.3 there exists a unique solution $\phi^*\in C_b^0([0,T]\times\mathbb{R}^d)$ which satisfies $T(t_0,\phi^*)=0$, namely φ_{ψ,t_0}. As we already established Fréchet differentiability of F in the first step, we are left to show invertibility of the operator $DF(t_0,\phi)[0,\Delta\phi]$. To this end, let $R:C_b^0([0,T]\times\mathbb{R}^d)\to C_b^0([0,T]\times\mathbb{R}^d)$ be the linear bounded operator defined by

$$(Rh)(\tau,x)=-h(\tau,x)+\int_\tau^{t_0}\int_{\mathbb{R}^d}h(s,y)\,\mathrm{d}[\eta(s,X_b(s;\tau,x))](y)g(s,\tau,x)\,\mathrm{d}s,$$

so that $DF(t_0,\phi)[0,\Delta\phi]=R\Delta\phi$. According to Inverse Mapping Theorem, it is sufficient to check that R is invertible which is equivalent to the existence and uniqueness of continuous solutions to the equation

$$(Rh)(\tau,x)=q(\tau,x)\qquad\forall q\in C_b^0([0,T]\times\mathbb{R}^d).$$

By Lemma 3.3.15 this equation is uniquely solvable, so that R and thus $DF(t_0,\phi)[0,\Delta\phi]$ are invertible. By the Implicit Function Theorem there exists some neighbourhood U of t_0 and a Fréchet differentiable function $\varphi(t)(\tau,x)$, so that

$$F(t,\varphi(t)(\tau,x))=0\qquad\forall t\in U.$$

On the other hand this equation is already uniquely solved by $\varphi_{\psi,t}=f(t,\cdot,\cdot)$, so that $t\mapsto\varphi_{\psi,t}(\tau,x)=f(t,\tau,x)$ is indeed Fréchet differentiable. $\square$

Now we are well prepared to show that μ_t is actually a weak solution in the sense of Definition 3.1.3. We already established nonnegativity (cf. Lemma 3.3.13) and narrow continuity (cf. Proposition 3.3.14), so that we are left to check the weak formulation.

Proposition 3.3.17. *The family of measures* $\{\mu_t\}_{t\in[0,T]}$ *be defined by* (3.3.26) *satisfies the weak formulation* (3.1.9) *with* $N = 0$.

Proof. Let $\varphi \in BL([0,T]\times\mathbb{R}^d)\cap C^1([0,T]\times\mathbb{R}^d)$ and consider the function

$$F : [0,T]\times[0,T] \to \mathbb{R}, \qquad F(s,t) := \int_{\mathbb{R}^d} \varphi(s,x)\,\mathrm{d}\mu_t(x).$$

Our goal is to compute the derivative of $\partial_\tau F(\tau,\tau)$ which will yield the claim. However, we first show that for all $\psi \in BL(\mathbb{R}^d)\cap C^1(\mathbb{R}^d)$ and all $t\in[0,T]$

$$\begin{aligned}\lim_{h\searrow 0}\frac{1}{h}&\int_{\mathbb{R}^d}\psi(x)\,\mathrm{d}[\mu_t-\mu_{t-h}](x)\\ &=\int_{\mathbb{R}^d}\left[b(t,x)\cdot\nabla_x\psi(x)+c(t,x)\psi(x)+\int_{\mathbb{R}^d}\psi(y)\,\mathrm{d}[\eta(t,x)](y)\right]\mathrm{d}\mu_t(x).\end{aligned} \tag{3.3.37}$$

Using (3.3.27) and the fact that $\varphi_{\psi,t}$ solves the dual problem (3.3.2), we see that for all $t\in[0,T]$

$$\begin{aligned}\frac{1}{h}\int_{\mathbb{R}^d}\psi(x)\,\mathrm{d}[\mu_t-\mu_{t-h}](x)&=\int_{\mathbb{R}^d}\frac{\varphi_{\psi,t}(t-h,x)-\varphi_{\psi,t}(t,x)}{h}\,\mathrm{d}\mu_{t-h}(x)\\ &\to-\int_{\mathbb{R}^d}\partial_\tau\varphi_{\psi,t}(\tau,x)\mid_{\tau=t}\mathrm{d}\mu_t(x)\qquad(h\to 0).\end{aligned} \tag{3.3.38}$$

More explicitly:

$$\begin{aligned}&\left|\int_{\mathbb{R}^d}\frac{\varphi_{\psi,t}(t,x)-\varphi_{\psi,t}(t-h,x)}{h}\,\mathrm{d}\mu_{t-h}(x)-\int_{\mathbb{R}^d}\partial_\tau\varphi_{\psi,t}(\tau,x)\mid_{\tau=t}\mathrm{d}\mu_t(x)\right|\\ \le&\left|\int_{\mathbb{R}^d}\partial_\tau\varphi_{\psi,t}(\tau,x)\mid_{\tau=t}\mathrm{d}[\mu_t-\mu_{t-h}](x)\right|\\ &+\left|\int_{\mathbb{R}^d}\partial_\tau\varphi_{\psi,t}(\tau,x)\mid_{\tau=t}-\frac{\varphi_{\psi,t}(t,x)-\varphi_{\psi,t}(t-h,x)}{h}\,\mathrm{d}\mu_{t-h}(x)\right|\\ =:&A+B.\end{aligned}$$

According to Proposition 3.3.7 $\varphi_{\psi,t}\in BL([0,T]\times\mathbb{R}^d)$ and consequently $\varphi_{\psi,t}$ is λ a.e. differentiable in τ according to Rademacher's Theorem and the derivative is a.e. bounded. However, we explicitly need differentiability for $\tau = t$. By construction $\varphi_{\psi,t}(t,\cdot) = \psi \in BL(\mathbb{R}^d) \cap C^1(\mathbb{R}^d)$ so that we can indeed conclude differentiability of $\varphi_{\psi,t}$ in $\tau = t$, and the derivative is even bounded. In particular, $\partial_\tau \varphi_{\psi,t}(\tau,\cdot) \mid_{\tau=t} \in C_b(\mathbb{R}^d)$ so that we conclude by narrow continuity $|A| \to 0$ for $h \to 0$.

Before we estimate term B, we note that according to Corollary 2.1.4 narrow continuity of the map $t \mapsto \mu_t$ implies that the family $\{\mu_{t-h} \mid h \geq 0\}$ is tight. Thus, for $\varepsilon > 0$ there exists a compact set $K \subset \mathbb{R}^d$ such that

$$\mu_{t-h}(K) \leq \varepsilon \qquad \forall h \geq 0.$$

Using K we split the integral of B

$$\begin{aligned} |B| \leq & \int_K \left| \partial_\tau \varphi_{\psi,t}(\tau,x) \mid_{\tau=t} - \frac{\varphi_{\psi,t}(t,x) - \varphi_{\psi,t}(t-h,x)}{h} \right| \mathrm{d}\mu_{t-h}(x) \\ & + \int_{\mathbb{R}^d \setminus K} \left| \partial_\tau \varphi_{\psi,t}(\tau,x) \mid_{\tau=t} - \frac{\varphi_{\psi,t}(t,x) - \varphi_{\psi,t}(t-h,x)}{h} \right| \mathrm{d}\mu_{t-h}(x). \end{aligned}$$

Remember that difference quotients converge uniformly to the derivative. Hence, for h small enough

$$\left| \partial_\tau \varphi_{\psi,t}(\tau,x) \mid_{\tau=t} - \frac{\varphi_{\psi,t}(t,x) - \varphi_{\psi,t}(t-h,x)}{h} \right| < 1 \qquad \text{for all } x \in \mathbb{R}^d$$

and we continue to estimate for small h

$$|B| \leq \left\| \partial_\tau \varphi_{\psi,t}(\tau,\cdot) \mid_{\tau=t} - \frac{\varphi_{\psi,t}(t,\cdot) - \varphi_{\psi,t}(t-h,\cdot)}{h} \right\|_{L^\infty(K)} \mu_{t-h}(K) + \mu_{t-h}(\mathbb{R}^d \setminus K).$$

Since $\mu_{t-h} \in \mathcal{M}^+(\mathbb{R}^d)$ for all h (see Lemma 3.3.13) we have with Proposition 3.3.14

$$\mu_{t-h}(K) \leq \mu_{t-h}(\mathbb{R}^d) \leq e^{t(\|c\|_\infty + \|\eta\|_\infty)} \|\mu_0\|_{TV} < \infty,$$

so that we conclude with the uniform convergence of the difference quotient and tightness

$$\limsup_{h \to 0} |B| \leq \varepsilon,$$

i.e. $|B| \to 0$ for $h \to 0$ proving (3.3.38).

With the formulation of the dual problem (3.3.2) we conclude (3.3.37) for all $t \in$

$[0,T]$

$$\lim_{h\searrow 0}\frac{1}{h}\int_{\mathbb{R}^d}\psi(x)\,\mathrm{d}[\mu_t-\mu_{t-h}](x) = -\int_{\mathbb{R}^d}\partial_\tau\varphi_{\psi,t}(\tau,x)\mid_{\tau=t}\mathrm{d}\mu_t(x)$$
$$=\int_{\mathbb{R}^d} b(t,x)\cdot\nabla_x\psi(x)+c(t,x)\psi(x)+\int_{\mathbb{R}^d}\psi(y)\,\mathrm{d}[\eta(t,x)](y)\,\mathrm{d}\mu_t(x). \tag{3.3.39}$$

We still need to prove a similar statement with the limit from above, i.e. that for all $t\in[0,T]$

$$\lim_{h\searrow 0}\frac{1}{h}\int_{\mathbb{R}^d}\psi(x)\,\mathrm{d}[\mu_{t+h}-\mu_t](x)$$
$$=\int_{\mathbb{R}^d}\left[b(t,x)\cdot\nabla_x\psi(x)+c(t,x)\psi(x)+\int_{\mathbb{R}^d}\psi(y)\,\mathrm{d}[\eta(t,x)](y)\right]\mathrm{d}\mu_t(x). \tag{3.3.40}$$

Note however, that $\varphi_{\psi,t}$ does not satisfy a semigroup property for $\tau>t$, so that we can not proceed as before. Instead, we rewrite the terms differently by applying (3.3.27) again which yields

$$\frac{1}{h}\int_{\mathbb{R}^d}\psi(x)\,\mathrm{d}[\mu_{t+h}-\mu_t](x)=\frac{1}{h}\int_{\mathbb{R}^d}\varphi_{\psi,t+h}(t+h,x)\,\mathrm{d}\mu_{t+h}(x)-\frac{1}{h}\int_{\mathbb{R}^d}\varphi_{\psi,t}(t,x)\,\mathrm{d}\mu_t(x)$$
$$=\int_{\mathbb{R}^d}\frac{\varphi_{\psi,t+h}(t,x)-\varphi_{\psi,t}(t,x)}{h}\,\mathrm{d}\mu_t(x). \tag{3.3.41}$$

According to Proposition 3.3.16 the map $t\mapsto\varphi_{\psi,t}(\tau,x)$ is Fréchet differentiable so that we compute with representation formula (3.3.3)

$$\begin{aligned}
&\partial_t\varphi_{\psi,t}(\tau,x)\\
=&\nabla_x\psi(X_b(t;\tau,x))\cdot b(t,X_b(t;\tau,x))e^{\int_\tau^t c(r,X_b(r;\tau,x))\,\mathrm{d}r}\\
&+\psi(X_b(t;\tau,x))e^{\int_\tau^t c(r,X_b(r;\tau,x))\,\mathrm{d}r}c(t,X_b(t;\tau,x))\\
&+\int_{\mathbb{R}^d}\varphi_{\psi,t}(t,y)\,\mathrm{d}[\eta(t,X_b(t;\tau,x))](y)e^{\int_\tau^t c(r,X_b(r;\tau,x))\,\mathrm{d}r}\\
&+\int_\tau^t\int_{\mathbb{R}^d}\partial_t\varphi_{\psi,t}(s,y)\,\mathrm{d}[\eta(s,X_b(s;\tau,x))](y)e^{\int_\tau^s c(r,X_b(r;\tau,x))\,\mathrm{d}r}\,\mathrm{d}s,
\end{aligned}$$

which implies

$$\partial_t\varphi_{\psi,t}(\tau,x)\mid_{(\tau,x)=(t,x)}=\nabla_x\psi(x)\cdot b(t,x)+\psi(x)c(t,x)+\int_{\mathbb{R}^d}\psi(y)\,\mathrm{d}[\eta(t,x)](y).$$

Plugging this into (3.3.41) we see

$$\begin{aligned}
&\frac{1}{h}\int_{\mathbb{R}^d} \psi(x)\,\mathrm{d}[\mu_{t+h}-\mu_t](x) = \int_{\mathbb{R}^d} \frac{\varphi_{\psi,t+h}(t,x)-\varphi_{\psi,t}(t,x)}{h}\,\mathrm{d}\mu_t(x)\\
&\to \int_{\mathbb{R}^d} \partial_t\varphi_{\psi,t}(\tau,x)\mid_{(\tau,x)=(t,x)}\,\mathrm{d}\mu_t(x) \qquad (3.3.42)\\
&= \int_{\mathbb{R}^d}\left[\nabla_x\psi(x)\cdot b(t,x)+\psi(x)c(t,x)+\int_{\mathbb{R}^d}\psi(y)\,\mathrm{d}[\eta(t,x)](y)\right]\mathrm{d}\mu_t(x),
\end{aligned}$$

proving (3.3.40). Combining (3.3.39) and (3.3.40) we conclude with Leibniz rule for all $\tau \in [0,T]$

$$\begin{aligned}
&\partial_\tau F(\tau,\tau)\\
&= \int_{\mathbb{R}^d}\left[\partial_\tau\varphi(\tau,x)+b(\tau,x)\cdot\nabla_x\varphi(\tau,x)+c(\tau,x)\varphi(\tau,x)+\int_{\mathbb{R}^d}\varphi(\tau,y)\,\mathrm{d}[\eta(\tau,x)](y)\right]\mathrm{d}\mu_\tau(x),
\end{aligned}$$

so integrating from $\tau = 0$ to $\tau = T$ yields (3.1.9) with $N = 0$. □

Remark 3.3.18. Note that if the model functions c, b, η are autonomous, i.e. time independent, it is not necessary to prove Fréchet differentiability in (3.3.41). In this case uniqueness of the dual problem (3.3.2) implies that for $h > 0$ and all $t \in [0, T-h]$

$$\varphi_{\psi,t+h}(\tau,x) = \varphi_{\psi,t}(\tau-h,x) \qquad \forall(\tau,x)\in[h,t+h]\times\mathbb{R}^d.$$

Consequently, we can rewrite (3.3.41) to

$$\begin{aligned}
\frac{1}{h}\int_{\mathbb{R}^d} \psi(x)\,\mathrm{d}[\mu_{t+h}-\mu_t](x) &= \int_{\mathbb{R}^d} \frac{\varphi_{\psi,t}(t-h,x)-\varphi_{\psi,t}(t,x)}{h}\,\mathrm{d}\mu_t(x)\\
&\to -\int_{\mathbb{R}^d}\partial_\tau\varphi_{\psi,t}(\tau,x)\mid_{\tau=t}\,\mathrm{d}\mu_t(x) \qquad (h\to 0),
\end{aligned}$$

where the limit $h \to 0$ can be computed analogously to term A in the first case. As we have already seen in (3.3.39) this implies (3.3.40).

Next, we show that the solutions to (3.3.1) are actually unique.

Lemma 3.3.19. *If* $\mu_0 = 0$*, then any measure solution* $\mu_\bullet$ *to* (3.3.1) *has to be the zero measure. In particular, solutions to* (3.3.1) *are unique.*

Proof. The idea of the proof follows [44, 3.39]. Let $\mu_\bullet$ be a measure solution to (3.3.1). As $\mu_0 = 0$ and $N = 0$ the weak formulation (3.1.9) reads for any test function $\varphi \in BL([0,T] \times \mathbb{R}^d) \cap C^1([0,T] \times \mathbb{R}^d)$

$$\begin{aligned}\int_{\mathbb{R}^d} \varphi(T,x)\,\mathrm{d}\mu_T(x) = &\int_0^T \int_{\mathbb{R}^d} \partial_t \varphi(t,x)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t \\ &+ \int_0^T \int_{\mathbb{R}^d} \big(\nabla_x \varphi(t,x) \cdot b(t,x) + \varphi(t,x)\,c(t,x)\big)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t \\ &+ \int_0^T \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \varphi(t,y)\,\mathrm{d}[\eta(t,x)]\,(y)\right) \mathrm{d}\mu_t(x)\,\mathrm{d}t,\end{aligned} \tag{3.3.43}$$

Now let $\psi \in BL([0,T] \times \mathbb{R}^d) \cap C^1([0,T] \times \mathbb{R}^d)$ be arbitrary and define $\varphi_{\psi,T}(t,x)$ as the unique solution of the dual problem

$$\begin{cases} \partial_\tau \varphi_{\psi,T} + b \cdot \nabla_x \varphi_{\psi,T} + c\,\varphi_{\psi,T} + \int_{\mathbb{R}^d} \varphi_{\psi,T}(t,y)\,\mathrm{d}[\eta(t,x)]\,(y) & = \psi \quad \text{in} \quad [0,T] \times \mathbb{R}^d, \\ \varphi_{\psi,T}(T,\cdot) & = 0 \quad \text{in} \quad \mathbb{R}^d. \end{cases} \tag{3.3.44}$$

Note the subtle differences to the dual problem (3.3.2) in the right hand side and the boundary condition. The solution $\varphi_{\psi,T}$ to (3.3.44) can be constructed similarly to the solution of (3.3.2) and has the same regularity $\varphi_{\psi,T} \in BL([0,T] \times \mathbb{R}^d)$. Testing (3.3.43) with $\varphi_{\psi,T}$ we obtain

$$\int_0^T \int_{\mathbb{R}^d} \psi(t,x)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t = 0 \qquad \forall \psi \in BL([0,T] \times \mathbb{R}^d) \cap C^1([0,T] \times \mathbb{R}^d). \tag{3.3.45}$$

To conclude that $\mu_\bullet$ is actually the zero measure, we restrict ourselves to test functions which are of the form $\psi(t,x) = \varphi(t)\xi(x)$ with $\varphi \in C_c^\infty([0,T])$ and $\xi \in C_c^\infty(\mathbb{R}^d)$. So instead of (3.3.45) we consider

$$\int_0^T \varphi(t) \int_{\mathbb{R}^d} \xi(x)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t = 0 \qquad \forall \varphi \in C_c^\infty([0,T]), \forall \xi \in C_c^\infty(\mathbb{R}^d). \tag{3.3.46}$$

From the Fundamental Lemma of Calculus of Variations and continuity of the map $t \mapsto \int_{\mathbb{R}^d} \xi(x)\,\mathrm{d}\mu_t(x)$ we obtain

$$\int_{\mathbb{R}^d} \xi(x)\,\mathrm{d}\mu_t(x) = 0 \qquad \forall \xi \in C_c^\infty(\mathbb{R}^d), \forall t \in [0,T], \tag{3.3.47}$$

i.e. that we can neglect the outer time integral. In this case, we conclude $\mu_t = 0$ for all $t \in [0,T]$ as $C_c^\infty(\mathbb{R}^d)$ is separating for $\mathcal{M}^+(\mathbb{R}^d)$ by Proposition 2.1.9. □

We want to conclude this section by proving continuity of solutions with respect to model functions and the initial measure. To this end, we first prove continuity of the solution to the dual problem (3.3.2).

Proposition 3.3.20. *Let $t \in [0,T]$ and let $(b,c,\eta), (\tilde{b},\tilde{c},\tilde{\eta})$ be two triples of model functions satisfying Assumption 3.1.1 with corresponding solutions $\varphi_{\psi,t}^{b,c,\eta}, \varphi_{\psi,t}^{\tilde{b},\tilde{c},\tilde{\eta}} \in BL([0,t]\times\mathbb{R}^d)$ to (3.3.2). Then*

$$\|\varphi_{\psi,t}^{b,c,\eta} - \varphi_{\psi,t}^{\tilde{b},\tilde{c},\tilde{\eta}}\|_\infty \leq t\max\{\|\psi\|_{BL},1\}(\|b-\tilde{b}\|_\infty + \|c-\tilde{c}\|_\infty + \|\eta-\tilde{\eta}\|_\infty)e^{Ct},$$

where C denotes a constant depending on the norms of all model functions.

Proof. We first note that for $f,\tilde{f} \in BL(\mathbb{R}^d)$ and $s,\tau \in [0,t]$

$$\begin{aligned} \|f(X_b(s;\tau,x)) - \tilde{f}(X_{\tilde{b}}(s;\tau,x))\| \leq & \|f-\tilde{f}\|_\infty + |f|_{\mathbf{Lip}}|X_b(s;\tau,x) - X_{\tilde{b}}(s;\tau,x)| \\ \leq & \|f-\tilde{f}\|_\infty + |f|_{\mathbf{Lip}}\|b-\tilde{b}\|_\infty t e^{2\|\tilde{b}\|_\infty t}, \end{aligned} \tag{3.3.48}$$

where we used Proposition 3.2.4.
Secondly, for $f,\tilde{f} \in BL([0,T]\times\mathbb{R}^d)$ we see with (3.3.48)

$$\begin{aligned} &\left| e^{\int_s^t f(u,X_b(u;s,x))\,\mathrm{d}u} - e^{\int_s^t \tilde{f}(u;X_{\tilde{b}}(u;s,x))\,\mathrm{d}u}\right| \\ \leq & e^{t(\|f\|_\infty + \|\tilde{f}\|_\infty)}\left|\int_s^t f(u;X_b(u;s,x)) - \tilde{f}(X_{\tilde{b}}(u;s,x))\,\mathrm{d}u\right| \\ \leq & t e^{t(\|f\|_\infty + \|\tilde{f}\|_\infty)}\left[\|f-\tilde{f}\|_\infty + |f|_{\mathbf{Lip}}\|b-\tilde{b}\|_\infty e^{2\|\tilde{b}\|_\infty t} t\right] \end{aligned} \tag{3.3.49}$$

where we used again that the exponential function $x \mapsto e^x$ is Lipschitz continuous on compact sets. Now we turn to the difference $\varphi_{\psi,t}^{b,c,\eta} - \varphi_{\psi,t}^{\tilde{b},\tilde{c},\tilde{\eta}}$. For $\lambda > 0$ to be chosen later we apply the usual Bielecki norm $\|\cdot\|_\lambda$ on $C^0([0,t]\times\mathbb{R}^d)$

$$\|g\|_\lambda = \sup_{(\tau,x)\in[0,t]\times\mathbb{R}^d} e^{-\lambda(t-\tau)}|g(\tau,x)| \tag{3.3.50}$$

Using (3.3.3) with triangle inequalities yields

$$
\begin{aligned}
&\|\varphi^{b,c,\eta}_{\psi,t} - \varphi^{\tilde b,\tilde c,\tilde\eta}_{\psi,t}\|_\lambda \\
\leq & \sup_{(\tau,x)\in[0,t]\times\mathbb{R}^d} e^{-\lambda(t-\tau)} \left|\psi(X_b(t;\tau,x)) - \psi(X_{\tilde b}(t;\tau,x))\right| e^{t\|c\|_\infty} \\
&+ \sup_{(\tau,x)\in[0,t]\times\mathbb{R}^d} e^{-\lambda(t-\tau)} \|\psi\|_\infty \left| e^{\int_\tau^t c(r,X_b(r;\tau,x))\,\mathrm{d}r} - e^{\int_\tau^t \tilde c(r,X_{\tilde b}(r;\tau,x))\,\mathrm{d}r}\right| \\
&+ \sup_{(\tau,x)\in[0,t]\times\mathbb{R}^d} e^{-\lambda(t-\tau)} \left|\int_\tau^t \int_{\mathbb{R}^d} \varphi^{b,c,\eta}_{\psi,t}(s,y) - \varphi^{\tilde b,\tilde c,\tilde\eta}_{\psi,t}(s,y)\,\mathrm{d}[\eta(s,X_b(s;\tau,x))](y)\,\mathrm{d}s\right| e^{t\|c\|_\infty} \\
&+ \sup_{(\tau,x)\in[0,t]\times\mathbb{R}^d} e^{-\lambda(t-\tau)} \left|\int_\tau^t \int_{\mathbb{R}^d} \varphi^{\tilde b,\tilde c,\tilde\eta}_{\psi,t}(s,y)\,\mathrm{d}[\eta(s,X_b(s;\tau,x)) - \tilde\eta(s,X_{\tilde b}(s;\tau,x))](y)\,\mathrm{d}s\right| e^{t\|c\|_\infty} \\
&+ \sup_{(\tau,x)\in[0,t]\times\mathbb{R}^d} e^{-\lambda(t-\tau)}\, t\, \|\varphi^{\tilde b,\tilde c,\tilde\eta}_{\psi,t}\|_\infty \|\tilde\eta\|_\infty \left| e^{\int_\tau^s c(r,X_b(r;\tau,x))\,\mathrm{d}r} - e^{\int_\tau^s \tilde c(r,X_{\tilde b}(r;\tau,x))\,\mathrm{d}r}\right|.
\end{aligned}
\tag{3.3.51}
$$

In the following, whenever we don't need the exponential factor of the Bielecki norm, we estimate it by 1. For the first two terms we notice with (3.3.48), (3.3.49) and Proposition 3.2.4

$$
\begin{aligned}
& \sup_{(\tau,x)\in[0,t]\times\mathbb{R}^d} \left|\psi(X_b(t;\tau,x)) - \psi(X_{\tilde b}(t;\tau,x))\right| e^{t\|c\|_\infty} \\
&+ \sup_{(\tau,x)\in[0,t]\times\mathbb{R}^d} \|\psi\|_\infty \left| e^{\int_\tau^t c(r,X_b(r;\tau,x))\,\mathrm{d}r} - e^{\int_\tau^t \tilde c(r,X_{\tilde b}(r;\tau,x))\,\mathrm{d}r}\right| \\
\leq & |\psi|_{\mathbf{Lip}} \|b-\tilde b\|_\infty t e^{t(2\|\tilde b\|_\infty + \|c\|_\infty)} + \|\psi\|_\infty t e^{t(\|c\|_\infty + \|\tilde c\|_\infty)} \left[\|c-\tilde c\|_\infty + \|c\|_\infty \|b-\tilde b\|_\infty e^{2\|\tilde b\|_\infty t} t\right] \\
\leq & t\|\psi\|_{BL} e^{t(2\|\tilde b\|_\infty + \|c\|_\infty + \|\tilde c\|_\infty)} \Big(\|b-\tilde b\|_\infty + \|c-\tilde c\|_\infty + \|c\|_\infty \|b-\tilde b\|_\infty t\Big).
\end{aligned}
\tag{3.3.52}
$$

The fourth term can be estimated by

$$
\begin{aligned}
& \sup_{(\tau,x)\in[0,t]\times\mathbb{R}^d} \left|\int_\tau^t \int_{\mathbb{R}^d} \varphi^{\tilde b,\tilde c,\tilde\eta}_{\psi,t}(s,y)\,\mathrm{d}[\eta(s,X_b(s;\tau,x)) - \tilde\eta(s,X_{\tilde b}(s;\tau,x))](y)\,\mathrm{d}s\right| e^{t\|c\|_\infty} \\
\leq & t\|\varphi^{\tilde b,\tilde c,\tilde\eta}_{\psi,t}\|_\infty \|\eta - \tilde\eta\|_\infty e^{t\|c\|_\infty}.
\end{aligned}
\tag{3.3.53}
$$

For the fifth term we apply (3.3.49) again

$$\begin{aligned}&\sup_{(\tau,x)\in[0,t]\times\mathbb{R}^d} t\|\varphi_{\psi,t}^{\tilde b,\tilde c,\tilde\eta}\|_\infty\|\tilde\eta\|_\infty \left|e^{\int_\tau^s c(r,X_b(r;\tau,x))\,\mathrm{d}r} - e^{\int_\tau^s \tilde c(r,X_{\tilde b}(r;\tau,x))\,\mathrm{d}r}\right|\\
\le& t\|\varphi_{\psi,t}^{\tilde b,\tilde c,\tilde\eta}\|_\infty\|\tilde\eta\|_\infty\left(te^{t(\|c\|_\infty+\|\tilde c\|_\infty)}\left[\|c-\tilde c\|_\infty+\|c\|_\infty\|b-\tilde b\|_\infty e^{2\|\tilde b\|_\infty t}t\right]\right)\\
\le& t^2\|\varphi_{\psi,t}^{\tilde b,\tilde c,\tilde\eta}\|_\infty\|\tilde\eta\|_\infty e^{t(2\|\tilde b\|_\infty+\|c\|_\infty+\|\tilde c\|_\infty)}\left[\|c-\tilde c\|_\infty+\|c\|_\infty\|b-\tilde b\|_\infty t\right].\end{aligned} \tag{3.3.54}$$

It remains to estimate the third term in (3.3.51).

$$\begin{aligned}&\sup_{(\tau,x)\in[0,t]\times\mathbb{R}^d} e^{-\lambda(t-\tau)}\left|\int_\tau^t\int_{\mathbb{R}^d}\varphi_{\psi,t}^{b,c,\eta}(s,y)-\varphi_{\psi,t}^{\tilde b,\tilde c,\tilde\eta}(s,y)\,\mathrm{d}[\eta(s,X_b(s;\tau,x))](y)\,\mathrm{d}s\right|e^{t\|c\|_\infty}\\
\le& e^{t\|c\|_\infty}\|\varphi_{\psi,t}^{b,c,\eta}-\varphi_{\psi,t}^{\tilde b,\tilde c,\tilde\eta}\|_\lambda\|\eta\|_\infty\sup_{\tau\in[0,t]}e^{-\lambda(t-\tau)}\int_\tau^t e^{\lambda(t-s)}\,\mathrm{d}s\\
=& e^{t\|c\|_\infty}\|\varphi_{\psi,t}^{b,c,\eta}-\varphi_{\psi,t}^{\tilde b,\tilde c,\tilde\eta}\|_\lambda\|\eta\|_\infty\sup_{\tau\in[0,T]}\frac{1-e^{-\lambda(t-\tau)}}{\lambda}\\
\le& e^{t\|c\|_\infty}\|\varphi_{\psi,t}^{b,c,\eta}-\varphi_{\psi,t}^{\tilde b,\tilde c,\tilde\eta}\|_\lambda\|\eta\|_\infty\frac{1-e^{-\lambda T}}{\lambda}\le\frac{e^{t\|c\|_\infty}\|\eta\|_\infty}{\lambda}\|\varphi_{\psi,t}^{b,c,\eta}-\varphi_{\psi,t}^{\tilde b,\tilde c,\tilde\eta}\|_\lambda.\end{aligned} \tag{3.3.55}$$

Choosing $\lambda = 2e^{t\|c\|_\infty}\|\eta\|_\infty$ and plugging (3.3.52), (3.3.53), (3.3.54) and (3.3.55) into (3.3.51) yields

$$\begin{aligned}&\|\varphi_{\psi,t}^{b,c,\eta}-\varphi_{\psi,t}^{\tilde b,\tilde c,\tilde\eta}\|_\lambda\\
\le& t\max\{\|\psi\|_{BL},1\}(\|b-\tilde b\|_\infty+\|c-\tilde c\|_\infty+\|\eta-\tilde\eta\|_\infty)e^{Ct}+\frac{1}{2}\|\varphi_{\psi,t}^{b,c,\eta}-\varphi_{\psi,t}^{\tilde b,\tilde c,\tilde\eta}\|_\lambda\end{aligned}$$

or equivalently

$$\|\varphi_{\psi,t}^{b,c,\eta}-\varphi_{\psi,t}^{\tilde b,\tilde c,\tilde\eta}\|_\lambda\le 2t\max\{\|\psi\|_{BL},1\}(\|b-\tilde b\|_\infty+\|c-\tilde c\|_\infty+\|\eta-\tilde\eta\|_\infty)e^{Ct}.$$

As $\|g\|_\lambda \ge e^{-\lambda t}\|g\|_\infty$ this yields the claim. □

Now we are able to prove continuity of μ_t with respect to intial measure and model functions.

Proposition 3.3.21. *Let the model functions b, c, η satisfy Assumption 3.1.1 and let $\mu_0 \in \mathcal{M}^+(\mathbb{R}^d)$ be any initial measure. Then the corresponding measure solution μ_t to* (3.3.1) *is Lipschitz continuous with respect to the initial measure and the model functions. More precisely,* *i) Let μ_t, ν_t be solutions to* (3.3.1) *with initial measures $\mu_0, \nu_0 \in \mathcal{M}^+(\mathbb{R}^d)$ respectively. Then*

$$\rho_F(\mu_t, \nu_t) \leq e^{Ct}\rho_F(\mu_0, \nu_0),$$

for some constant $C > 0$ depending on the model functions b, c, η but independent on μ_0, ν_0.

ii) Let $(b, c, \eta), (\tilde{b}, \tilde{c}, \tilde{\eta})$ be two triples of model functions, satisfying Assumption 3.1.1, and let $\mu_t^{b,c,\eta}, \mu_t^{\tilde{b},\tilde{c},\tilde{\eta}}$ be the corresponding solutions to (3.3.1) *with initial measure μ_0. Then*

$$\rho_F\left(\mu_t^{b,c,\eta}, \mu_t^{\tilde{b},\tilde{c},\tilde{\eta}}\right) \leq t(\|b-\tilde{b}\|_\infty + \|c-\tilde{c}\|_\infty + \|\eta-\tilde{\eta}\|_\infty)e^{Ct}\|\mu_0\|_{TV},$$

where C denotes a constant depending on the norms of all model functions.

Proof. i) Using (3.3.26) and Proposition 3.3.4 we compute

$$\begin{aligned}\rho_F(\mu_t, \nu_t) &= \sup_{\|\psi\|_{BL\leq 1}} \int_{\mathbb{R}^d} \psi(x)\,\mathrm{d}(\mu_t-\nu_t)(x) = \sup_{\|\psi\|_{BL\leq 1}} \int_{\mathbb{R}^d} \varphi_{\psi,t}(0,x)\,\mathrm{d}(\mu_0-\nu_0)(x)\\ &\leq \sup_{\|\psi\|_{BL\leq 1}} \|\varphi_{\psi,t}\|_\infty \int_{\mathbb{R}^d} 1\,\mathrm{d}(\mu_0-\nu_0)(x) \leq e^{Ct}\rho_F(\mu_0,\nu_0).\end{aligned}$$

ii) Similarly, using (3.3.26) and Proposition 3.3.20 we have an estimate of the form

$$\begin{aligned}\rho_F(\mu_t^{b,c,\eta}, \mu_t^{\tilde{b},\tilde{c},\tilde{\eta}}) &= \sup_{\|\psi\|_{BL\leq 1}} \int_{\mathbb{R}^d} \psi(x)\,\mathrm{d}\left(\mu_t^{b,c,\eta} - \mu_t^{\tilde{b},\tilde{c},\tilde{\eta}}\right)(x)\\ &= \sup_{\|\psi\|_{BL\leq 1}} \int_{\mathbb{R}^d} \varphi_{\psi,t}^{b,c,\eta}(0,x) - \varphi_{\psi,t}^{\tilde{b},\tilde{c},\tilde{\eta}}(0,x)\,\mathrm{d}\mu_0(x)\\ &\leq \sup_{\|\psi\|_{BL\leq 1}} \|\varphi_{\psi,t}^{b,c,\eta}(0,\cdot) - \varphi_{\psi,t}^{\tilde{b},\tilde{c},\tilde{\eta}}(0,\cdot)\|_\infty \|\mu_0\|_{TV}\\ &\leq t(\|b-\tilde{b}\|_\infty + \|c-\tilde{c}\|_\infty + \|\eta-\tilde{\eta}\|_\infty)e^{Ct}\|\mu_0\|_{TV}.\end{aligned}$$

□

3.4 Adding state-independent influx to the model

Now we consider the full linear model (3.1.1) with state-independent influx N. This case is a bit trickier since we can not go to the dual problem as the primal problem is not linear in μ_t. Instead we will invoke Duhamel's principle to construct a solution. In particular, let $\varphi_{\psi,t}$ be the solution to the dual problem (3.3.2), i.e. the dual problem to (3.1.1) with $N = 0$. Then plugging $\varphi_{\psi,t}$ into (3.1.9) for $T = t$ yields for the left-hand side

$$\int_{\mathbb{R}^d} \varphi_{\psi,t}(\tau, x)\,\mathrm{d}\mu_t(x) - \int_{\mathbb{R}^d} \varphi_{\psi,t}(0, x)\,\mathrm{d}\mu_0(x) = \int_{\mathbb{R}^d} \psi(x)\,\mathrm{d}\mu_t(x) - \int_{\mathbb{R}^d} \varphi_{\psi,t}(0, x)\,\mathrm{d}\mu_0(x)$$

and for the right-hand side, using that $\varphi_{\psi,t}$ solves (3.3.2),

$$\begin{aligned}
&\int_0^t \int_{\mathbb{R}^d} (\partial_\tau \varphi_{\psi,t}(\tau, x) + \nabla_x \varphi_{\psi,t}(\tau, x) \cdot b(\tau, x) + \varphi_{\psi,t}(\tau, x)\, c(\tau, x))\,\mathrm{d}\mu_\tau(x)\,\mathrm{d}\tau \\
&\qquad + \int_0^t \left(\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \varphi_{\psi,t}(\tau, y)\,\mathrm{d}[\eta(\tau, x)]\,(y)\,\mathrm{d}\mu_\tau(x) + \int_{\mathbb{R}^d} \varphi_{\psi,t}(\tau, x)\,\mathrm{d}[N(\tau)](x) \right) \mathrm{d}\tau \\
&\quad = \int_0^t \int_{\mathbb{R}^d} \varphi_{\psi,t}(\tau, x)\,\mathrm{d}[N(\tau)](x)\,\mathrm{d}\tau.
\end{aligned} \tag{3.4.1}$$

So in accordance with (3.3.26) we define for all $t \in [0, T]$

$$\int_{\mathbb{R}^d} \psi(x)\,\mathrm{d}\mu_t(x) = \int_{\mathbb{R}^d} \varphi_{\psi,t}(0, x)\,\mathrm{d}\mu_0(x) + \int_0^t \int_{\mathbb{R}^d} \varphi_{\psi,t}(\tau, x)\,\mathrm{d}[N(\tau)](x)\,\mathrm{d}\tau. \tag{3.4.2}$$

Lemma 3.4.1. *Let μ_t be defined by* (3.4.2)*. Then for all $\psi \in BL(\mathbb{R}^d)$ and all $0 \le s \le t \le T$*

$$\int_{\mathbb{R}^d} \psi(x)\,\mathrm{d}\mu_t(x) = \int_{\mathbb{R}^d} \varphi_{\psi,t}(s, x)\,\mathrm{d}\mu_s(x) + \int_s^t \int_{\mathbb{R}^d} \varphi_{\psi,t}(\tau, x)\,\mathrm{d}[N(\tau)](x)\,\mathrm{d}\tau. \tag{3.4.3}$$

Proof. Let $s \in [0, t]$ and let $\xi(\tau, x)$ be the solution to the dual problem (3.3.2) terminating at time s with value $\xi(s, x) = \varphi_{\psi,t}(s, x)$. We directly conclude $\xi(\tau, x) = \varphi_{\psi,t}(\tau, x)$ for all $\tau \in [0, s]$ by uniqueness of the dual problem. Applying (3.4.2) twice yields

$$\int_{\mathbb{R}^d} \psi(x)\,\mathrm{d}\mu_t(x) = \int_{\mathbb{R}^d} \varphi_{\psi,t}(0,x)\,\mathrm{d}\mu_0(x) + \int_0^t \int_{\mathbb{R}^d} \varphi_{\psi,t}(\tau,x)\,\mathrm{d}[N(\tau)](x)\,\mathrm{d}\tau$$
$$= \int_{\mathbb{R}^d} \xi(0,x)\,\mathrm{d}\mu_0(x) + \int_0^s \int_{\mathbb{R}^d} \xi(\tau,x)\,\mathrm{d}[N(\tau)](x)\,\mathrm{d}\tau + \int_s^t \int_{\mathbb{R}^d} \varphi_{\psi,t}(\tau,x)\,\mathrm{d}[N(\tau)](x)\,\mathrm{d}\tau$$
$$= \int_{\mathbb{R}^d} \varphi_{\psi,t}(s,x)\,\mathrm{d}\mu_s(x) + \int_s^t \int_{\mathbb{R}^d} \varphi_{\psi,t}(\tau,x)\,\mathrm{d}[N(\tau)](x)\,\mathrm{d}\tau,$$

as desired. □

Lemma 3.4.2. *Let μ_t be defined by (3.4.2). Then for all $\psi \in BL(\mathbb{R}^d)$ and all $t \in [0,T]$ the measure $\mu_t \in \mathcal{M}^+(\mathbb{R}^d)$.*

Proof. In the proof of Lemma 3.3.13 we showed that $\psi \geq 0$ implies $\varphi_{\psi,t} \geq 0$. As both μ_0 and $N(t)$ are in $\mathcal{M}^+(\mathbb{R}^d)$ for all $t \in [0,T]$, we conclude by (3.4.2). □

Similarly to Proposition 3.3.14 we can prove:

Proposition 3.4.3. *Let μ_t be definded by (3.4.2). Then the following statements hold.*

i) μ_t is bounded in total variation norm, i.e.

$$\|\mu_t\|_{TV} \leq e^{t(\|c\|_\infty + \|\eta\|_\infty)}(\|\mu_0\|_{TV} + \|N\|_\infty t).$$

ii) The map $t \mapsto \mu_t$ is narrowly continuous. Furthermore, the map $t \mapsto \mu_t$ is Lipschitz continuous.

Proof. The proof of i) is similar to the proof corresponding to Proposition 3.3.14. Just note that the extra term including N can be bounded with Proposition 3.3.4 by

$$\sup_{\substack{\psi \in C_c^1(\mathbb{R}^d) \\ \|\psi\|_\infty \leq 1}} \int_0^t \int_{\mathbb{R}^d} \varphi_{\psi,t}(\tau,x)\,\mathrm{d}[N(\tau)](x)\,\mathrm{d}\tau \leq e^{t(\|c\|_\infty + \|\eta\|_\infty)}\|N\|_\infty t,$$

so that

$$\|\mu_t\|_{TV} \leq e^{t(\|c\|_\infty + \|\eta\|_\infty)}(\|\mu_0\|_{TV} + \|N\|_\infty t).$$

To show the narrow continuity, analogously to the proof of Proposition 3.3.14 ii) we prove that for all $\psi \in BL(\mathbb{R}^d)$ the map $t \mapsto \int_{\mathbb{R}^d} \psi\,\mathrm{d}\mu_t$ is Lipschitz continuous. This

then implies narrow continuity by Portemanteau Theorem.
To check the Lipschitz continuity let $s_1 > s_2$, then with Lemmata 3.4.1 and 3.3.9

$$\begin{aligned}
&\left|\int_{\mathbb{R}^d} \psi(x)\,\mathrm{d}[\mu_{s_1} - \mu_{s_2}](x)\right| \\
=&\left|\int_{\mathbb{R}^d} \varphi_{\psi,s_1}(s_2,x)\,\mathrm{d}\mu_{s_2}(x) + \int_{s_2}^{s_1}\int_{\mathbb{R}^d} \varphi_{\psi,s_1}(\tau,x)\,\mathrm{d}[N(\tau)](x)\,\mathrm{d}\tau - \int_{\mathbb{R}^d} \varphi_{\psi,s_2}(s_2,x)\,\mathrm{d}\mu_{s_2}(x)\right| \\
\leq&\int_{\mathbb{R}^d} |\varphi_{\psi,s_1}(s_2,x) - \varphi_{\psi,s_2}(s_2,x)|\,\mathrm{d}\mu_{s_2}(x) + \left|\int_{s_2}^{s_1}\int_{\mathbb{R}^d} \varphi_{\psi,s_1}(\tau,x)\,\mathrm{d}[N(\tau)](x)\,\mathrm{d}\tau\right| \\
\leq& C|s_1 - s_2|\|\psi\|_{BL}\|\mu_{s_2}\|_{BL^*} + \|\varphi_{\psi,s_1}\|_{BL}\|N\|_\infty|s_1 - s_2|.
\end{aligned}$$

So using i) and Proposition 3.3.4 we get the desired Lipschitz continuity

$$\left|\int_{\mathbb{R}^d} \psi(x)\,\mathrm{d}[\mu_{s_1} - \mu_{s_2}](x)\right| \leq C\|\psi\|_{BL}|s_1 - s_2|.$$

As in the proof of Proposition 3.3.14, this implies Lipschitz continuity of the map $t \mapsto \mu_t$, i.e.

$$\rho_F(\mu_{s_1}, \mu_{s_2}) = \sup_{\|\psi\|_{BL}\leq 1} \int_{\mathbb{R}^d} \psi(x)\,\mathrm{d}[\mu_{s_1} - \mu_{s_2}](x) \leq C|s_1 - s_2|.$$

□

Now, similarly to Proposition 3.3.17 we can show that μ_t is actually a solution to (3.1.1) in the sense of Definition 3.1.3.

Proposition 3.4.4. *The family of measures* $\{\mu_t\}_{t\in[0,T]}$ *be defined by* (3.4.2) *satisfies the weak formulation* (3.1.9).

Proof. We adjust the proof of Proposition 3.3.17 to account for the additional term. Let $\varphi \in BL([0,T] \times \mathbb{R}^d)$ and consider the function

$$f : [0,T] \times [0,T] \to \mathbb{R}, \qquad f(s,t) := \int_{\mathbb{R}^d} \varphi(s,x)\,\mathrm{d}\mu_t(x).$$

We first show that for all $\psi \in BL(\mathbb{R}^d)$ and a.e. $t \in [0,T]$

$$\begin{aligned}
&\lim_{h\searrow 0} \frac{1}{h} \int_{\mathbb{R}^d} \psi(x)\,\mathrm{d}[\mu_t - \mu_{t-h}](x) \\
&= \int_{\mathbb{R}^d} \left[b(t,x)\cdot \nabla_x \psi(x) + c(t,x)\psi(x) + \int_{\mathbb{R}^d} \psi(y)\,\mathrm{d}[\eta(t,x)](y) \right] \mathrm{d}\mu_t(x) \\
&\quad + \int_{\mathbb{R}^d} \psi(x)\,\mathrm{d}[N(t)](x).
\end{aligned} \tag{3.4.4}$$

Indeed, using the fact that $\varphi_{\psi,t}$ solves the dual problem (3.3.2) and Lemma 3.4.1, we see that for a.e. $t \in [0,T]$

$$\begin{aligned}
&\frac{1}{h} \int_{\mathbb{R}^d} \psi(x)\,\mathrm{d}[\mu_t - \mu_{t-h}](x) \\
&= \int_{\mathbb{R}^d} \frac{\varphi_{\psi,t}(t-h,x) - \varphi_{\psi,t}(t,x)}{h}\,\mathrm{d}\mu_{t-h}(x) + \frac{1}{h}\int_{t-h}^{t} \int_{\mathbb{R}^d} \varphi_{\psi,t}(\tau,x)\,\mathrm{d}[N(\tau)](x)\,\mathrm{d}\tau.
\end{aligned} \tag{3.4.5}$$

According to the proof of Proposition 3.3.17 we already know that

$$\begin{aligned}
&\lim_{h\searrow 0} \int_{\mathbb{R}^d} \frac{\varphi_{\psi,t}(t-h,x) - \varphi_{\psi,t}(t,x)}{h}\,\mathrm{d}\mu_{t-h}(x) \\
&= \int_{\mathbb{R}^d} \left[b(t,x)\cdot \nabla_x \psi(x) + c(t,x)\psi(x) + \int_{\mathbb{R}^d} \psi(y)\,\mathrm{d}[\eta(t,x)](y) \right] \mathrm{d}\mu_t(x),
\end{aligned}$$

so we only have to consider the second term in (3.4.5). By Assumptions 3.1.1 and Proposition 3.3.7 we see that the map $\tau \mapsto \int_{\mathbb{R}^d} \varphi_{\psi,t}(\tau,x)\,\mathrm{d}[N(\tau)](x)$ is continuous, so that we can apply Lebesgue's Differentiation Theorem which yields

$$\begin{aligned}
&\lim_{h\searrow 0} \frac{1}{h}\int_{t-h}^{t} \int_{\mathbb{R}^d} \varphi_{\psi,t}(\tau,x)\,\mathrm{d}[N(\tau)](x)\,\mathrm{d}\tau \\
&= \int_{\mathbb{R}^d} \varphi_{\psi,t}(t,x)\,\mathrm{d}[N(t)](x) = \int_{\mathbb{R}^d} \psi(x)\,\mathrm{d}[N(t)](x),
\end{aligned} \tag{3.4.6}$$

proving (3.4.4). Analogously, we show

$$\begin{aligned}
&\lim_{h\searrow 0} \frac{1}{h} \int_{\mathbb{R}^d} \psi(x)\,\mathrm{d}[\mu_{t+h} - \mu_t](x) \\
&= \int_{\mathbb{R}^d} \left[b(t,x)\cdot \nabla_x \psi(x) + c(t,x)\psi(x) + \int_{\mathbb{R}^d} \psi(y)\,\mathrm{d}[\eta(t,x)](y) \right] \mathrm{d}\mu_t(x) \\
&\quad + \int_{\mathbb{R}^d} \psi(x)\,\mathrm{d}[N(t)](x).
\end{aligned} \tag{3.4.7}$$

To this end, again applying Lemma 3.4.1 we see

$$\begin{aligned}&\frac{1}{h}\int_{\mathbb{R}^d}\psi(x)\,\mathrm{d}[\mu_{t+h}-\mu_t](x)\\=&\frac{1}{h}\int_{\mathbb{R}^d}\varphi_{\psi,t+h}(t+h,x)\,\mathrm{d}\mu_{t+h}(x)-\frac{1}{h}\int_{\mathbb{R}^d}\varphi_{\psi,t}(t,x)\,\mathrm{d}\mu_t(x)\\=&\int_{\mathbb{R}^d}\frac{\varphi_{\psi,t+h}(t,x)-\varphi_{\psi,t}(t,x)}{h}\,\mathrm{d}\mu_t(x)+\frac{1}{h}\int_t^{t+h}\int_{\mathbb{R}^d}\varphi_{\psi,t+h}(\tau,x)\,\mathrm{d}[N(\tau)](x)\,\mathrm{d}\tau.\end{aligned}\tag{3.4.8}$$

The first term has already been treated in the proof of Proposition 3.3.17, see (3.3.42), which yields

$$\begin{aligned}&\int_{\mathbb{R}^d}\frac{\varphi_{\psi,t+h}(t,x)-\varphi_{\psi,t}(t,x)}{h}\,\mathrm{d}\mu_t(x)\to\int_{\mathbb{R}^d}\partial_t\varphi_{\psi,t}(\tau,x)\mid_{(\tau,x)=(t,x)}\,\mathrm{d}\mu_t(x)\\=&\int_{\mathbb{R}^d}\left[\nabla_x\psi(x)\cdot b(t,x)+\psi(x)c(t,x)+\int_{\mathbb{R}^d}\psi(y)\,\mathrm{d}[\eta(t,x)](y)\right]\mathrm{d}\mu_t(x).\end{aligned}\tag{3.4.9}$$

We turn to the second term in (3.4.8).

$$\begin{aligned}&\frac{1}{h}\int_t^{t+h}\int_{\mathbb{R}^d}\varphi_{\psi,t+h}(\tau,x)\,\mathrm{d}[N(\tau)](x)\,\mathrm{d}\tau\\=&\frac{1}{h}\int_t^{t+h}\int_{\mathbb{R}^d}\varphi_{\psi,t}(\tau,x)\,\mathrm{d}[N(\tau)](x)\,\mathrm{d}\tau\\&+\int_t^{t+h}\int_{\mathbb{R}^d}\frac{\varphi_{\psi,t+h}(\tau,x)-\varphi_{\psi,t}(\tau,x)}{h}\,\mathrm{d}[N(\tau)](x)\,\mathrm{d}\tau.\end{aligned}\tag{3.4.10}$$

The first term can be handled similarly to (3.4.6) with Lebesgue's Differentiation Theorem yielding

$$\lim_{h\searrow 0}\frac{1}{h}\int_t^{t+h}\int_{\mathbb{R}^d}\varphi_{\psi,t}(\tau,x)\,\mathrm{d}[N(\tau)](x)\,\mathrm{d}\tau=\int_{\mathbb{R}^d}\psi(x)\,\mathrm{d}[N(t)](x).\tag{3.4.11}$$

According to Proposition 3.3.16 the map $t\mapsto\varphi_{\psi,t}$ is Fréchet differentiable with $\frac{\varphi_{\psi,t+h}-\varphi_{\psi,t}}{h}$ converging uniformly to the derivative. In particular, $\frac{\varphi_{\psi,t+h}(\tau,x)-\varphi_{\psi,t}(\tau,x)}{h}$ is uniformly bounded for small h and thus with Dominated Convergence Theorem

$$\int_t^{t+h}\int_{\mathbb{R}^d}\frac{\varphi_{\psi,t+h}(\tau,x)-\varphi_{\psi,t}(\tau,x)}{h}\,\mathrm{d}[N(\tau)](x)\,\mathrm{d}\tau\to 0\qquad(h\to 0).\tag{3.4.12}$$

Hence, plugging (3.4.11) and (3.4.12) into (3.4.10) yields

$$\frac{1}{h}\int_t^{t+h}\int_{\mathbb{R}^d}\varphi_{\psi,t+h}(\tau,x)\,\mathrm{d}[N(\tau)](x)\,\mathrm{d}\tau \to \int_{\mathbb{R}^d}\psi(x)\,\mathrm{d}[N(t)](x) \qquad (h\to 0)$$

which together with (3.4.9) in (3.4.8) leads to

$$\begin{aligned}\frac{1}{h}\int_{\mathbb{R}^d}\psi(x)\,\mathrm{d}[\mu_{t+h}-\mu_t](x) \to &\int_{\mathbb{R}^d}\nabla_x\psi(x)\cdot b(t,x)+\psi(x)c(t,x)\,\mathrm{d}\mu_t(x)\\ &+\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\psi(y)\,\mathrm{d}[\eta(t,x)](y)\,\mathrm{d}\mu_t(x)+\int_{\mathbb{R}^d}\psi(x)\,\mathrm{d}[N(t)](x) \qquad (h\to 0)\end{aligned}$$

proving (3.4.7).

We combine (3.4.4) and (3.4.7) and conclude with Leibniz rule for a.e. $\tau\in[0,T]$

$$\begin{aligned}\partial_\tau f(\tau,\tau)=&\int_{\mathbb{R}^d}[\partial_\tau\varphi(\tau,x)+b(\tau,x)\cdot\nabla_x\varphi(\tau,x)+c(\tau,x)\varphi(\tau,x)]\,\mathrm{d}\mu_\tau(x)\\ &+\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\varphi(\tau,y)\,\mathrm{d}[\eta(\tau,x)](y)\,\mathrm{d}\mu_\tau(x)+\int_{\mathbb{R}^d}\varphi(\tau,x)\,\mathrm{d}[N(\tau)](x),\end{aligned}$$

so integrating from $\tau=0$ to $\tau=T$ yields (3.1.9). ☐

The uniqueness result Lemma 3.3.19 of the previous section with $N=0$ also implies uniqueness of solutions to the full linear model (3.1.1).

Lemma 3.4.5. *Under Assumptions 3.1.1 solutions to* (3.1.1) *are unique.*

Proof. Let $\mu_\bullet,\tilde{\mu}_\bullet$ be two solutions to (3.1.1) with initial measure μ_0. Then a direct computation reveals that $\mu_\bullet-\tilde{\mu}_\bullet$ solves (3.3.1) with initial measure $\tilde{\mu}_0=0$. According to Lemma 3.3.19 $\mu_\bullet-\tilde{\mu}_\bullet=0$ which yields the claim. ☐

As a last step, we generalise the continuity result Proposition 3.3.21 to the case with state-independent influx.

Proposition 3.4.6. *Let the model functions b,c,η,N satisfy Assumption 3.1.1 and let $\mu_0\in\mathcal{M}^+(\mathbb{R}^d)$ be any initial measure. Then the corresponding measure solution μ_t to* (3.1.1) *is Lipschitz continuous with respect to the initial measure and the model functions. More precisely,*

i) Let μ_t,ν_t be solutions to (3.1.1) *with initial measures $\mu_0,\nu_0\in\mathcal{M}^+(\mathbb{R}^d)$ respectively. Then*

$$\rho_F(\mu_t,\nu_t)\le e^{Ct}\rho_F(\mu_0,\nu_0),$$

for some constant $C > 0$ depending on the norms of model functions b, c, η but independent on μ_0, ν_0.

ii) Let $(b, c, \eta, N), (\tilde{b}, \tilde{c}, \tilde{\eta}, \tilde{N})$ be two vectors of model functions, satisfying Assumption 3.1.1, and let $\mu_t^{b,c,\eta,N}, \mu_t^{\tilde{b},\tilde{c},\tilde{\eta},\tilde{N}}$ be the corresponding solutions to (3.1.1) with initial measure μ_0. Then

$$\rho_F\left(\mu_t^{b,c,\eta,N}, \mu_t^{\tilde{b},\tilde{c},\tilde{\eta},\tilde{N}}\right) \le te^{Ct}\left(\|b-\tilde{b}\|_\infty + \|c-\tilde{c}\|_\infty + \|\eta-\tilde{\eta}\|_\infty + \|N-\tilde{N}\|_\infty\right),$$

where C denotes a constant depending on norms of all model functions and μ_0.

Proof. We first show i). A direct computation with (3.4.2) yields

$$\begin{aligned}
\rho_F(\mu_t, \nu_t) &= \sup_{\|\psi\|_{BL}\le 1} \int_{\mathbb{R}^d} \psi(x)\,\mathrm{d}(\mu_t - \nu_t)(x) \\
&= \sup_{\|\psi\|_{BL}\le 1} \bigg[\int_{\mathbb{R}^d} \varphi_{\psi,t}(0,x)\,\mathrm{d}\mu_0(x) + \int_0^t \int_{\mathbb{R}^d} \varphi_{\psi,t}(\tau,x)\,\mathrm{d}[N(\tau)](x)\,\mathrm{d}\tau \\
&\quad - \int_{\mathbb{R}^d} \varphi_{\psi,t}(0,x)\,\mathrm{d}\nu_0(x) - \int_0^t \int_{\mathbb{R}^d} \varphi_{\psi,t}(\tau,x)\,\mathrm{d}[N(\tau)](x)\,\mathrm{d}\tau\bigg] \\
&= \int_{\mathbb{R}^d} \varphi_{\psi,t}(0,x)\,\mathrm{d}(\mu_0 - \nu_0)(x).
\end{aligned}$$

So similarly to the proof of Proposition 3.3.21 we apply Proposition 3.3.4 to see

$$\rho_F(\mu_t, \nu_t) \le \sup_{\|\psi\|_{BL\le 1}} \|\varphi_{\psi,t}\|_\infty \int_{\mathbb{R}^d} 1\,\mathrm{d}(\mu_0 - \nu_0)(x) \le e^{Ct}\rho_F(\mu_0, \nu_0).$$

Before we show ii) in a similar computation as before, note that by construction the solution $\varphi_{\psi,t}$ to the dual problem (3.3.2) only depends on the model functions

(b, c, η). Hence, we compute with (3.4.2)

$$
\begin{aligned}
&\rho_F(\mu_t^{b,c,\eta,N}, \mu_t^{\tilde{b},\tilde{c},\tilde{\eta},\tilde{N}}) = \sup_{\|\psi\|_{BL\le 1}} \int_{\mathbb{R}^d} \psi(x)\,\mathrm{d}\Big(\mu_t^{b,c,\eta,N} - \mu_t^{\tilde{b},\tilde{c},\tilde{\eta},\tilde{N}}\Big)(x) \\
&= \sup_{\|\psi\|_{BL\le 1}} \int_{\mathbb{R}^d} \varphi_{\psi,t}^{b,c,\eta}(0,x) - \varphi_{\psi,t}^{\tilde{b},\tilde{c},\tilde{\eta}}(0,x)\,\mathrm{d}\mu_0(x) \\
&\quad + \int_0^t \int_{\mathbb{R}^d} \varphi_{\psi,t}^{b,c,\eta}(\tau,x)\,\mathrm{d}[N(\tau)](x)\,\mathrm{d}\tau - \int_0^t \int_{\mathbb{R}^d} \varphi_{\psi,t}^{\tilde{b},\tilde{c},\tilde{\eta}}(\tau,x)\,\mathrm{d}[\tilde{N}(\tau)](x)\,\mathrm{d}\tau \\
&\le \sup_{\|\psi\|_{BL\le 1}} \|\varphi_{\psi,t}^{b,c,\eta}(0,\cdot) - \varphi_{\psi,t}^{\tilde{b},\tilde{c},\tilde{\eta}}(0,\cdot)\|_\infty \|\mu_0\|_{TV} \\
&\quad + \sup_{\|\psi\|_{BL\le 1}} \int_0^t \int_{\mathbb{R}^d} \varphi_{\psi,t}^{b,c,\eta}(\tau,x) - \varphi_{\psi,t}^{\tilde{b},\tilde{c},\tilde{\eta}}(\tau,x)\,\mathrm{d}[N(\tau)](x)\,\mathrm{d}\tau \\
&\quad + \sup_{\|\psi\|_{BL\le 1}} \int_0^t \int_{\mathbb{R}^d} \varphi_{\psi,t}^{\tilde{b},\tilde{c},\tilde{\eta}}(\tau,x)\,\mathrm{d}\Big[N(\tau) - \tilde{N}(\tau)\Big](x)\,\mathrm{d}\tau
\end{aligned}
$$

We apply Proposition 3.3.20 to bound the first two terms and use Proposition 3.3.4 to estimate the third term, i.e.

$$
\begin{aligned}
&\rho_F(\mu_t^{b,c,\eta,N}, \mu_t^{\tilde{b},\tilde{c},\tilde{\eta},\tilde{N}}) \\
&\le t(\|b-\tilde{b}\|_\infty + \|c-\tilde{c}\|_\infty + \|\eta-\tilde{\eta}\|_\infty)e^{Ct}\|\mu_0\|_{TV} \\
&\quad + \sup_{\|\psi\|_{BL\le 1}} \|\varphi_{\psi,t}^{b,c,\eta} - \varphi_{\psi,t}^{\tilde{b},\tilde{c},\tilde{\eta}}\|_\infty \|N\|_\infty t + \sup_{\|\psi\|_{BL}\le 1} \|\varphi_{\psi,t}^{\tilde{b},\tilde{c},\tilde{\eta}}\|_\infty \sup_{\tau\in[0,T]} \|N(\tau) - \tilde{N}(\tau)\|_{BL^*}\, t \\
&\le te^{Ct}\Big(\|b-\tilde{b}\|_\infty + \|c-\tilde{c}\|_\infty + \|\eta-\tilde{\eta}\|_\infty + \|N-\tilde{N}\|_\infty\Big).
\end{aligned}
$$

□

We conclude the treatment of the linear model (3.1.1) with a theorem combining the results of this section.

Theorem 3.4.7. *Under Assumptions 3.1.1 there exists a unique Lipschitz continuous measure solution $\mu : [0,T] \to \mathcal{M}^+(\mathbb{R}^d)$ to model* (3.1.1) *in the sense of Definition 3.1.3. The solution $\mu_\bullet$ is Lipschitz continuous with respect to model functions and the initial measure μ_0.*

3.5 The nonlinear case

Now we construct a solution to the nonlinear version of (3.1.1) on $[0,T]\times\mathbb{R}^d$, i.e.

$$\begin{cases} \partial_t\mu_t + \nabla_x\cdot(b(t,x,\mu_t)\mu_t) & = c(t,x,\mu_t)\mu_t + \int_{\mathbb{R}^d}\eta(t,x,\mu_t)(y)\,\mathrm{d}\mu_t(y) + N(t,\mu_t), \\ \mu_0 & = \nu, \end{cases} \tag{3.5.1}$$

where $\nu\in\mathcal{M}^+(\mathbb{R}^d)$ and

$$\begin{aligned} &b:[0,T]\times\mathbb{R}^d\times\mathcal{M}^+(\mathbb{R}^d)\to\mathbb{R}^d, && c:[0,T]\times\mathbb{R}^d\times\mathcal{M}^+(\mathbb{R}^d)\to\mathbb{R}, \\ &\eta:[0,T]\times\mathbb{R}^d\times\mathcal{M}^+(\mathbb{R}^d)\to\mathcal{M}^+(\mathbb{R}^d), && N:[0,T]\times\mathcal{M}^+(\mathbb{R}^d)\to\mathcal{M}^+(\mathbb{R}^d). \end{aligned}$$

For any $\varphi\in BL([0,T]\times\mathbb{R}^d)\cap C^1([0,T]\times\mathbb{R}^d)$ the corresponding **weak formulation** is given by

$$\begin{aligned} &\int_{\mathbb{R}^d}\varphi(T,x)\,\mathrm{d}\mu_T(x) - \int_{\mathbb{R}^d}\varphi(0,x)\,\mathrm{d}\mu_0(x) = \int_0^T\int_{\mathbb{R}^d}\partial_t\varphi(t,x)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t \\ &\quad + \int_0^T\int_{\mathbb{R}^d}\big(\nabla_x\varphi(t,x)\cdot b(t,x,\mu_t) + \varphi(t,x)\,c(t,x,\mu_t)\big)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t \\ &\quad + \int_0^T\int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d}\varphi(t,y)\,\mathrm{d}[\eta(t,x,\mu_t)]\,(y)\right)\mathrm{d}\mu_t(x)\,\mathrm{d}t + \int_0^T\int_{\mathbb{R}^d}\varphi(t,x)\,\mathrm{d}[N(t,\mu_t)](x)\,\mathrm{d}t. \end{aligned} \tag{3.5.2}$$

We need the following regularity of our model functions.

Assumptions 3.5.1 (Nonlinear model)**.** The model functions b,c,η and N should satisfy the following conditions.

(i) For any $\mu\in\mathcal{M}^+(\mathbb{R}^d)$, the model functions $b(\cdot,\cdot,\mu), c(\cdot,\cdot,\mu), \eta(\cdot,\cdot,\mu)$ and $N(\cdot,\mu)$ satisfy Assumption 3.1.1. Moreover, we have uniform bounds

$$\int_0^T\sup_{\mu\in\mathcal{M}^+(\mathbb{R}^d)}\Big[\|c(\tau,\cdot,\mu)\|_\infty + \sup_{y\in\mathbb{R}^d}\|\eta(\tau,y,\mu)\|_{BL^*} + \|N(\tau,\mu)\|_{BL^*}\Big]\,\mathrm{d}\tau<\infty,$$

and for any $R>0$

$$\int_0^T\sup_{\|\mu\|_{BL^*}\le R}\Big[\|b(\tau,\cdot,\mu)\|_{BL} + |c(\tau,\cdot,\mu)|_{\mathbf{Lip}} + \|\eta(\tau,\cdot,\mu)\|_{BL(\mathbb{R}^d;\mathcal{M}^+)}\Big]\,\mathrm{d}\tau<\infty.$$

(ii) For any $R>0$, there exists a nonnegative function $L_{R,b}$ in $C^0(0,T)$ so that if

$\mu_\bullet$, $\nu_\bullet$ are narrowly continuous and $\|\mu\|_{BL^*}, \|\nu\|_{BL^*} \leq R$, then

$$\|b(t,\cdot,\mu) - b(t,\cdot,\nu)\|_\infty \leq L_{R,b}(t)\rho_F(\mu,\nu).$$

(iii) For any $R > 0$, there exists a nonnegative function $L_{R,c} \in C^0(0,T)$ so that for $\|\mu\|_{BL^*}, \|\nu\|_{BL^*} \leq R$

$$\|c(t,\cdot,\mu) - c(t,\cdot,\nu)\|_\infty \leq L_{R,c}(t)\rho_F(\mu,\nu).$$

(iv) For any $R > 0$, there exists a nonnegative function $L_{R,\eta} \in C^0(0,T)$ so that for $\|\mu\|_{BL^*}, \|\nu\|_{BL^*} \leq R$

$$\sup_{y\in\mathbb{R}^d} \rho_F\left(\eta(t,y,\mu), \eta(t,y,\nu)\right) \leq L_{R,\eta}(t)\,\rho_F(\mu,\nu).$$

(v) For any $R > 0$, there exists a nonnegative function $L_{R,N} \in C^0(0,T)$ so that if $\|\mu\|_{BL^*}, \|\nu\|_{BL^*} \leq R$

$$\sup_{y\in\mathbb{R}^d} \rho_F\left(N(t,\mu), N(t,\nu)\right) \leq L_{R,N}(t)\,\rho_F(\mu,\nu).$$

In this section we will construct solutions to (3.5.1) by reducing the nonlinear model to a linear model. To this end, we split the time interval into subintervals on which we freeze the arguments of the nonlinearities. More precisely, fix $k \in \mathbb{N}$ and divide $[0,T]$ into 2^k intervals of the form $[lT/2^k, (l+1)T/2^k]$, $l = 0, ..., 2^k - 1$. For $t \in [lT/2^k, (l+1)T/2^k]$ we define the **k-th approximation** μ_t^k recursively as solution to the *linear* problem

$$\begin{aligned}&\partial_t\mu_t + \nabla_x\cdot\left(b\left(t,x,\mu_{\frac{lT}{2^k}}\right)\mu_t\right)\\ &\quad= c\left(t,x,\mu_{\frac{lT}{2^k}}\right)\mu_t + \int_{\mathbb{R}^d}\eta\left(t,x,\mu_{\frac{lT}{2^k}}\right)(y)\,\mathrm{d}\mu_t(y) + N\left(t,\mu_{\frac{lT}{2^k}}\right)\end{aligned} \tag{3.5.3}$$

with initial condition $\mu_t|_{t=lT/2^k} = \mu_{lT/2^k}$ which we obtain from solving the corresponding problem on the previous interval $\left[(l-1)T/2^k, lT/2^k\right]$. Naturally, we set $\mu_t\mid_{t=0}= \mu_0$.

We will show that $\left(\mu_t^k\right)_{k\in\mathbb{N}}$ defines a Cauchy sequence in $C^0([0,T],\mathcal{M}^+(\mathbb{R}^d))$ which converges to the solution of (3.5.1). To this end, we first need the following simple lemma.

Lemma 3.5.2. *[44, 2.22] Let a and b be arbitrary positive constants. If a sequence $(u_l)_{l\in\mathbb{N}_0}\subset\mathbb{R}$ satisfies the inequality,*

$$|u_{l+1}|\leq a|u_l|+b,\quad l=0,1,2,\dots,\tag{3.5.4}$$

Then, it holds

$$|u_l|\leq a^l|u_0|+\begin{cases}\frac{a^l-1}{a-1}b & \text{for } a\neq 1,\\ lb & \text{for } a=1.\end{cases}\tag{3.5.5}$$

Now we turn to study the sequence $\left(\mu_t^k\right)_{k\in\mathbb{N}}$ and notice that it is uniformly bounded in $\mathcal{M}^+(\mathbb{R}^d)$.

Lemma 3.5.3. *The family of measures $\{\mu_t^k\}_{t\in[0,T]}^k$ satisfies*

$$\|\mu_t^k\|_{TV}\leq(\|\mu_0\|_{TV}+2\|N\|_\infty T)e^{2CT}$$

for some constant $C>0$ depending on norms of model functions but independent of k.

Proof. Let $k\in\mathbb{N}$, $l\in\{0,...,2^k\}$ and observe that according to Proposition 3.4.3 we have for $t\in[lT/2^k,(l+1)T/2^k]$

$$\begin{aligned}\|\mu_t^k\|_{TV}&\leq\left[\|\mu^k_{\frac{lT}{2^k}}\|_{TV}+\|N\|_\infty\left(\frac{(l+1)T}{2^k}-\frac{lT}{2^k}\right)\right]e^{CT\left(\frac{l+1}{2^k}-\frac{l}{2^k}\right)}\\&=\left[\|\mu^k_{\frac{lT}{2^k}}\|_{TV}+\|N\|_\infty\frac{T}{2^k}\right]e^{\frac{CT}{2^k}}.\end{aligned}\tag{3.5.6}$$

Applying Lemma 3.5.2 for $t=\frac{(l+1)T}{2^k}$ to $u_l:=\|\mu^k_{\frac{lT}{2^k}}\|_{TV}$ yields

$$\|\mu^k_{\frac{lT}{2^k}}\|_{TV}\leq e^{\frac{CTl}{2^k}}\|\mu_0\|_{TV}+\frac{e^{\frac{CTl}{2^k}}-1}{e^{\frac{CT}{2^k}}-1}e^{\frac{CT}{2^k}}\|N\|_\infty\frac{T}{2^k}.\tag{3.5.7}$$

We further estimate inequality (3.5.7) by noticing that for $x>1$ we have $x^l-1=$

$(x-1)\sum_{i=0}^{l-1} x^i$ so that

$$\frac{x^l-1}{x-1}x = \sum_{i=1}^{l} x^i \leq l x^l.$$

Thus, we get with $x = e^{\frac{CT}{2^k}}$

$$\|\mu^k_{\frac{lT}{2^k}}\|_{TV} \leq e^{CT}\|\mu_0\|_{TV} + \|N\|_\infty l e^{\frac{CTl}{2^k}}\frac{T}{2^k} \leq e^{CT}\left[\|\mu_0\|_{TV} + \|N\|_\infty T\right], \tag{3.5.8}$$

which implies the assertion for $t = lT/2^k$, $l = 0,...,2^k$. Applying (3.5.8) to (3.5.6) yields the claim for all $t \in [0,T]$

$$\begin{aligned}\|\mu^k_t\|_{TV} &\leq \left[\|\mu^k_{\frac{lT}{2^k}}\|_{TV} + \|N\|_\infty \frac{T}{2^k}\right] e^{\frac{CT}{2^k}} \leq \left[e^{CT}\left[\|\mu_0\|_{TV} + \|N\|_\infty T\right] + \|N\|_\infty \frac{T}{2^k}\right] e^{CT} \\ &\leq e^{2CT}\left[\|\mu_0\|_{TV} + 2\|N\|_\infty T\right].\end{aligned}$$

□

Lemma 3.5.4. *There exists a constant $C >$ depending on norms of model functions but independent of k such that we have*

$$\rho_F\left(\mu^{k+1}_t, \mu^k_t\right) \leq C 2^{-k}.$$

Proof. We adjust to proof of [44, 2.24] to our purposes.
Let $k \in \mathbb{N}$ and $l \in \{0,...,2^k-1\}$. Furthermore, let $t \in [lT/2^k, (l+1)T/2^k] =: A_l$. As the interval A_l gets split in the iteration step $k \mapsto k+1$ we denote by A_l^- and A_l^+ the left and right half of A_l respectively. In particular, we introduce

$$t_* := \frac{lT}{2^k} \qquad \text{and} \qquad t_m := t_* + \frac{T}{2^{k+1}},$$

so that $A_l^- = [t_*, t_m]$. As a first step, let $t \in A_l^-$. We introduce an intermediate measure $\tilde{\mu}^k_t$ which starts at t_* with initial datum $\mu^k_{t_*}$ similarly to μ^k_t but evolves with the nonlinearities of μ^{k+1}_t, i.e. the nonlinearities are evaluated at $\mu^{k+1}_{t_*}$. To facilitate understanding of the following calculations, we have summarised the various measures here once again

Measure	initial value at t_*	evolves with model functions
μ_t^k	$\mu_{t_*}^k$	$b(\cdot,\cdot,\mu_{t_*}^k), c(\cdot,\cdot,\mu_{t_*}^k), \eta(\cdot,\cdot,\mu_{t_*}^k), N(\cdot,\mu_{t_*}^k)$
$\tilde{\mu}_{t_k}^k$	$\mu_{t_*}^k$	$b(\cdot,\cdot,\mu_{t_*}^{k+1}), c(\cdot,\cdot,\mu_{t_*}^{k+1}), \eta(\cdot,\cdot,\mu_{t_*}^{k+1}), N(\cdot,\mu_{t_*}^{k+1})$
$\mu_{t_k}^{k+1}$	$\mu_{t_*}^{k+1}$	$b(\cdot,\cdot,\mu_{t_*}^{k+1}), c(\cdot,\cdot,\mu_{t_*}^{k+1}), \eta(\cdot,\cdot,\mu_{t_*}^{k+1}), N(\cdot,\mu_{t_*}^{k+1})$

This allows us to apply triangle inequality as well as the continuity results of Proposition 3.4.6 to get

$$\begin{aligned}
&\rho_F\left(\mu_t^{k+1},\mu_t^k\right) \leq \rho_F\left(\mu_t^{k+1},\tilde{\mu}_t^k\right) + \rho_F\left(\tilde{\mu}_t^k,\mu_t^k\right)\\
\leq& Ce^{C(t-t_*)}\rho_F\left(\mu_{t_*}^{k+1},\mu_{t_*}^k\right)\\
&+(t-t_*)e^{C(t-t_*)}\left(\|b\left(\cdot,\cdot,\mu_{t_*}^k\right)-b\left(\cdot,\cdot,\mu_{t_*}^{k+1}\right)\|_\infty + \|c\left(\cdot,\cdot,\mu_{t_*}^k\right)-c\left(\cdot,\cdot,\mu_{t_*}^{k+1}\right)\|_\infty\right)\\
&+(t-t_*)e^{C(t-t_*)}\left(\|\eta\left(\cdot,\cdot,\mu_{t_*}^k\right)-\eta\left(\cdot,\cdot,\mu_{t_*}^{k+1}\right)\|_\infty + \|N\left(\cdot,\mu_{t_*}^k\right)-N\left(\cdot,\mu_{t_*}^{k+1}\right)\|_\infty\right).
\end{aligned}$$

According to Lemma 3.5.3 $\|\mu_t^k\|_{TV}$ is uniformly bounded by some $R>0$ for all k and with the same lemma also $\|\tilde{\mu}_t^k\|_{TV}$. Hence, with Assumptions 3.5.1 (ii)-(v) we estimate

$$\begin{aligned}
&\rho_F\left(\mu_t^{k+1},\mu_t^k\right)\\
\leq& e^{C2^{-(k+1)}}\rho_F\left(\mu_{t_*}^{k+1},\mu_{t_*}^k\right)\\
&+T2^{-(k+1)}e^{C2^{-(k+1)}}\rho_F\left(\mu_{t_*}^{k+1},\mu_{t_*}^k\right)\sup_{\tau\in[0,T]}\left[L_{R,b}(\tau)+L_{R,c}(\tau)+L_{R,\eta}(\tau)+L_{R,N}(\tau)\right]\\
\leq& e^{C2^{-(k+1)}}\rho_F\left(\mu_{t_*}^{k+1},\mu_{t_*}^k\right) + C2^{-(k+1)}e^{C2^{-(k+1)}}\rho_F\left(\mu_{t_*}^{k+1},\mu_{t_*}^k\right)\\
\leq& e^{C2^{-(k+1)}}\rho_F\left(\mu_{t_*}^{k+1},\mu_{t_*}^k\right).
\end{aligned} \tag{3.5.9}$$

Now let $t\in A_l^+$. Similarly to before, we introduce an intermediate measure $\tilde{\mu}_t^k$ but this time starting at t_m with initial measure $\mu_{t_m}^k$ but evolving with nonlinearities evaluated at $\mu_{t_m}^{k+1}$.

Measure	initial value at t_m	evolves with model functions
μ_t^k	$\mu_{t_m}^k$	$b(\cdot,\cdot,\mu_{t_*}^k), c(\cdot,\cdot,\mu_{t_*}^k), \eta(\cdot,\cdot,\mu_{t_*}^k), N(\cdot,\mu_{t_*}^k)$
$\tilde{\mu}_{t_k}^k$	$\mu_{t_m}^k$	$b(\cdot,\cdot,\mu_{t_m}^{k+1}), c(\cdot,\cdot,\mu_{t_m}^{k+1}), \eta(\cdot,\cdot,\mu_{t_m}^{k+1}), N(\cdot,\mu_{t_m}^{k+1})$
$\mu_{t_k}^{k+1}$	$\mu_{t_m}^{k+1}$	$b(\cdot,\cdot,\mu_{t_m}^{k+1}), c(\cdot,\cdot,\mu_{t_m}^{k+1}), \eta(\cdot,\cdot,\mu_{t_m}^{k+1}), N(\cdot,\mu_{t_*}^{k+1})$

Similar to the case $t \in A_l^-$ we apply triangle inequality as well as the continuity results of Proposition 3.4.6 to see

$$\begin{aligned}
&\rho_F\left(\mu_t^{k+1},\mu_t^k\right) \leq \rho_F\left(\mu_t^{k+1},\tilde{\mu}_t^k\right) + \rho_F\left(\tilde{\mu}_t^k,\mu_t^k\right)\\
\leq & e^{C(t-t_*)}\rho_F\left(\mu_{t_m}^{k+1},\mu_{t_m}^k\right)\\
&+(t-t_*)e^{C(t-t_*)}\left(\|b\left(\cdot,\cdot,\mu_{t_*}^k\right)-b\left(\cdot,\cdot,\mu_{t_m}^{k+1}\right)\|_\infty + \|c\left(\cdot,\cdot,\mu_{t_*}^k\right)-c\left(\cdot,\cdot,\mu_{t_m}^{k+1}\right)\|_\infty\right)\\
&+(t-t_*)e^{C(t-t_*)}\left(\|\eta\left(\cdot,\cdot,\mu_{t_*}^k\right)-\eta\left(\cdot,\cdot,\mu_{t_m}^{k+1}\right)\|_\infty + \|N\left(\cdot,\mu_{t_*}^k\right)-N\left(\cdot,\mu_{t_m}^{k+1}\right)\|_\infty\right).
\end{aligned}$$

We again apply Assumptions 3.5.1 (ii)-(v) and see

$$\begin{aligned}
&\rho_F\left(\mu_t^{k+1},\mu_t^k\right)\\
\leq & e^{C2^{-(k+1)}}\rho_F\left(\mu_{t_m}^{k+1},\mu_{t_m}^k\right)\\
&+T2^{-(k+1)}e^{C2^{-(k+1)}}\rho_F\left(\mu_{t_*}^k,\mu_{t_m}^{k+1}\right)\sup_{\tau\in[0,T]}\left[L_{R,b}(\tau)+L_{R,c}(\tau)+L_{R,\eta}(\tau)+L_{R,N}(\tau)\right]\\
\leq & e^{C2^{-(k+1)}}\rho_F\left(\mu_{t_m}^{k+1},\mu_{t_m}^k\right) + C2^{-(k+1)}e^{C2^{-(k+1)}}\rho_F\left(\mu_{t_*}^k,\mu_{t_m}^{k+1}\right).
\end{aligned} \tag{3.5.10}$$

With triangle inequality, Lipschitz continuity in time (see Proposition 3.4.3) and (3.5.9) for $t = t_m \in A_l^-$ we see that

$$\begin{aligned}
\rho_F\left(\mu_{t_*}^k,\mu_{t_m}^{k+1}\right) \leq & \rho_F\left(\mu_{t_*}^k,\mu_{t_m}^k\right) + \rho_F\left(\mu_{t_m}^k,\mu_{t_m}^{k+1}\right) \leq C|t_m-t_*| + e^{C2^{-(k+1)}}\rho_F\left(\mu_{t_*}^{k+1},\mu_{t_*}^k\right)\\
\leq & C2^{-(k+1)} + e^{C2^{-(k+1)}}\rho_F\left(\mu_{t_*}^{k+1},\mu_{t_*}^k\right).
\end{aligned}$$

We insert this estimate into (3.5.10) and apply (3.5.9) again which leads to

$$\begin{aligned}
&\rho_F\left(\mu_t^{k+1}, \mu_t^k\right) \\
\leq & e^{C2^{-(k+1)}} \rho_F\left(\mu_{t_m}^{k+1}, \mu_{t_m}^k\right) + C2^{-(k+1)} e^{C2^{-(k+1)}} \left[C2^{-(k+1)} + e^{C2^{-(k+1)}} \rho_F\left(\mu_{t_*}^{k+1}, \mu_{t_*}^k\right)\right] \\
\leq & e^{C2^{-(k+1)}} \rho_F\left(\mu_{t_*}^{k+1}, \mu_{t_*}^k\right) + C\left(2^{-(k+1)}\right)^2 e^{C2^{-(k+1)}}
\end{aligned} \tag{3.5.11}$$

for all $t \in A_l^+$. Combining (3.5.9) and (3.5.11) yields that for all $t \in A_l$

$$\rho_F\left(\mu_t^{k+1}, \mu_t^k\right) \leq e^{C2^{-(k+1)}} \rho_F\left(\mu_{t_*}^{k+1}, \mu_{t_*}^k\right) + C\left(2^{-(k+1)}\right)^2 e^{C2^{-(k+1)}}. \tag{3.5.12}$$

Next, we set $t = (m+1)T/2^k$ and apply Lemma 3.5.2 to the sequence $u_m := \rho_F\left(\mu_{mT/2^k}^{k+1}, \mu_{mT/2^k}^k\right)$ to deduce

$$\begin{aligned}
u_m \leq & e^{mC2^{-(k+1)}} \rho_F\left(\mu_0^{k+1}, \mu_0^k\right) + \frac{e^{mC2^{-(k+1)}} - 1}{e^{C2^{-(k+1)}} - 1} C\left(2^{-(k+1)}\right)^2 e^{C2^{-(k+1)}} \\
= & 2^{-(k+1)} \frac{C2^{-(k+1)}}{e^{C2^{-(k+1)}} - 1} \left(e^{mC2^{-(k+1)}} - 1\right) e^{C2^{-(k+1)}} \\
\leq & 2^{-(k+1)} (e^C - 1) e^C.
\end{aligned} \tag{3.5.13}$$

In the last step, we used $x + 1 \leq e^x$, $m \leq 2^k$ and that the limit $\lim_{x \to 0} x/(e^x - 1)$ exists. This yields the claim for all t of the form $mT/2^k$. For all other t we invoke (3.5.12) together with (3.5.13) to see

$$\begin{aligned}
\rho_F\left(\mu_t^{k+1}, \mu_t^k\right) \leq & e^{C2^{-(k+1)}} \rho_F\left(\mu_{t_*}^{k+1}, \mu_{t_*}^k\right) + C\left(2^{-(k+1)}\right)^2 e^{C2^{-(k+1)}} \\
\leq & e^{C2^{-(k+1)}} 2^{-(k+1)} (e^C - 1) e^C + C\left(2^{-(k+1)}\right)^2 e^{C2^{-(k+1)}} \leq C2^{-(k+1)}
\end{aligned}$$

as desired. □

Now with the above results, we are able to prove the existence of solutions to the nonlinear model.

Theorem 3.5.5. *Under Assumptions 3.5.1 there exists a Lipschitz continuous solution $\mu_\bullet : [0,T] \to \mathcal{M}^+(\mathbb{R}^d)$ to (3.5.1). The solution depends Lipschitz continuously on the initial measure as well as on the model functions.*

Proof. Let $n \in \mathbb{N}$. For $m > n$ we compute, using Lemma 3.5.4

$$\rho_F(\mu_t^m, \mu_t^n) \leq \rho_F(\mu_t^m, \mu_t^{m-1}) + \ldots + \rho_F(\mu_t^{n+1}, \mu_t^n) \leq C\left(2^{-(m-1)} + \ldots + 2^{-n}\right) \leq C2 \cdot 2^{-n}.$$

where C is the constant from Lemma 3.5.4 which is independent of $t \in [0,T]$. Thus, $(\mu_t^k)_{k\in\mathbb{N}}$ is a Cauchy sequence in $C^0\left([0,T];\mathcal{M}^+(\mathbb{R}^d)\right)$ and as such, there is $\mu_t \in C^0\left([0,T];\mathcal{M}^+(\mathbb{R}^d)\right)$ such that

$$\mu_t^k \to \mu_t \text{ in } C^0\left([0,T];\mathcal{M}^+(\mathbb{R}^d)\right).$$

By construction μ_t satisfies the same bound in $\|\cdot\|_{TV}$ as μ_t^k which is given in Lemma 3.5.3. Therefore, with Assumptions 3.5.1 (ii) we have

$$\|b\left(\cdot,\cdot,\mu_t^k\right) - b\left(\cdot,\cdot,\mu_t\right)\|_\infty \leq \sup_{\tau\in[0,T]} L_{R,b}(\tau)\,\rho_F(\mu_t^k,\mu_t) \leq C\rho_F\left(\mu_t^k,\mu_t\right) \tag{3.5.14}$$

and similar estimates for the other model functions c, η, N. Thus, we can pass to the limit in the weak formulation of (3.5.3) proving that μ_t satisfies the weak formulation (3.5.2). More precisely, let $\varphi \in BL([0,T]\times\mathbb{R}^d) \cap C^1([0,T]\times\mathbb{R}^d)$ and consider the RHS of (3.5.2)

$$\begin{aligned}
&\int_0^T \int_{\mathbb{R}^d} \partial_t\varphi(t,x)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t \\
&+ \int_0^T \int_{\mathbb{R}^d} \big(\nabla_x\varphi(t,x)\cdot b(t,x,\mu_t) + \varphi(t,x)\,c(t,x,\mu_t)\big)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t \\
&+ \int_0^T \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \varphi(t,y)\,\mathrm{d}[\eta(t,x,\mu_t)]\,(y)\right)\mathrm{d}\mu_t(x)\,\mathrm{d}t + \int_0^T \int_{\mathbb{R}^d} \varphi(t,x)\,\mathrm{d}[N(t,\mu_t)](x)\,\mathrm{d}t \\
=:& A + B + \mathcal{C} + D.
\end{aligned}$$

We introduce the approximation μ_t^k and compute

$$\begin{aligned}
A =& \int_0^T \int_{\mathbb{R}^d} \partial_t\varphi(t,x)\,\mathrm{d}[\mu_t - \mu_t^k](x)\,\mathrm{d}t + \int_0^T \int_{\mathbb{R}^d} \partial_t\varphi(t,x)\,\mathrm{d}\mu_t^k(x)\,\mathrm{d}t \\
\leq& \|\varphi\|_{BL} \int_0^T \rho_F(\mu_t,\mu_t^k)\,\mathrm{d}t + \int_0^T \int_{\mathbb{R}^d} \partial_t\varphi(t,x)\,\mathrm{d}\mu_t^k(x)\,\mathrm{d}t \\
\leq& C \sup_{t\in[0,T]} \rho_F(\mu_t,\mu_t^k) + \int_0^T \int_{\mathbb{R}^d} \partial_t\varphi(t,x)\,\mathrm{d}\mu_t^k(x)\,\mathrm{d}t.
\end{aligned}$$

In order to rewrite term B we remember that $t \mapsto \mu_t$ is continuous so that $\sup_{t\in[0,T]}\|\mu_t\|_{BL^*}$

is bounded and use the estimates of the form (3.5.14).

$$
\begin{aligned}
&B \\
=&\int_0^T \int_{\mathbb{R}^d} \left(\nabla_x\varphi(t,x)\cdot(b(t,x,\mu_t)-b(t,x,\mu_t^k))+\varphi(t,x)\left(c(t,x,\mu_t)-c(t,x,\mu_t^k\right)\mathrm{d}\mu_t(x)\,\mathrm{d}t \\
&+\int_0^T \int_{\mathbb{R}^d} \left(\nabla_x\varphi(t,x)\cdot b(t,x,\mu_t^k)+\varphi(t,x)\,c(t,x,\mu_t^k)\right)\mathrm{d}[\mu_t-\mu_t^k](x)\,\mathrm{d}t \\
&+\int_0^T \int_{\mathbb{R}^d} \left(\nabla_x\varphi(t,x)\cdot b(t,x,\mu_t^k)+\varphi(t,x)\,c(t,x,\mu_t^k)\right)\mathrm{d}\mu_t^k(x)\,\mathrm{d}t \\
\leq&\|\varphi\|_{BL}\left[\|b(\cdot,\cdot,\mu_t)-b(\cdot,\cdot,\mu_t^k)\|_\infty+\|c(\cdot,\cdot,\mu_t)-c(\cdot,\cdot,\mu_t^k)\|_\infty\right]\sup_{t\in[0,T]}\|\mu_t\|_{BL^*}T \\
&+\|\varphi\|_{BL}(\|b\|_\infty+\|c\|_\infty)\int_0^T \rho_F(\mu_t,\mu_t^k)\,\mathrm{d}t \\
&+\int_0^T \int_{\mathbb{R}^d} \left(\nabla_x\varphi(t,x)\cdot b(t,x,\mu_t^k)+\varphi(t,x)\,c(t,x,\mu_t^k)\right)\mathrm{d}\mu_t^k(x)\,\mathrm{d}t \\
\leq&C\sup_{t\in[0,T]}\rho_F(\mu_t,\mu_t^k)+\int_0^T \int_{\mathbb{R}^d} \left(\nabla_x\varphi(t,x)\cdot b(t,x,\mu_t^k)+\varphi(t,x)\,c(t,x,\mu_t^k)\right)\mathrm{d}\mu_t^k(x)\,\mathrm{d}t.
\end{aligned}
$$

Similarly, we can bound the terms $\mathcal{C}$ and D by

- $\mathcal{C} \leq C\sup_{t\in[0,T]}\rho_F(\mu_t,\mu_t^k)+\int_0^T \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d}\varphi(t,y)\,\mathrm{d}\big[\eta(t,x,\mu_t^k)\big]\,(y)\right)\mathrm{d}\mu_t^k(x)\,\mathrm{d}t,$
- $D \leq C\sup_{t\in[0,T]}\rho_F(\mu_t,\mu_t^k)+\int_0^T \int_{\mathbb{R}^d} \varphi(t,x)\,\mathrm{d}[N(t,\mu_t^k)](x)\,\mathrm{d}t.$

Combinig the bounds on $A, B, \mathcal{C}, D$ with the fact that μ_t^k solves the weak formulation leads to

$$
A+B+\mathcal{C}+D\leq C\sup_{t\in[0,T]}\rho_F(\mu_t,\mu_t^k)+\int_{\mathbb{R}^d}\varphi(T,x)\,\mathrm{d}\mu_T^k(x)-\int_{\mathbb{R}^d}\varphi(0,x)\,\mathrm{d}\mu_0^k(x).
$$

Next, we use that $\mu_0^k=\mu_0$ and another triangle inequality to see

$$
\begin{aligned}
&A+B+\mathcal{C}+D \\
\leq&C\sup_{t\in[0,T]}\rho_F(\mu_t,\mu_t^k)+\int_{\mathbb{R}^d}\varphi(T,x)\,\mathrm{d}\mu_T(x)-\int_{\mathbb{R}^d}\varphi(0,x)\,\mathrm{d}\mu_0(x)+\int_{\mathbb{R}^d}\varphi(T,x)\,\mathrm{d}[\mu_T^k-\mu_T](x) \\
\leq&C\sup_{t\in[0,T]}\rho_F(\mu_t,\mu_t^k)+\int_{\mathbb{R}^d}\varphi(T,x)\,\mathrm{d}\mu_T(x)-\int_{\mathbb{R}^d}\varphi(0,x)\,\mathrm{d}\mu_0(x)+C\rho_F(\mu_T,\mu_T^k) \\
\leq&C\sup_{t\in[0,T]}\rho_F(\mu_t,\mu_t^k)+\int_{\mathbb{R}^d}\varphi(T,x)\,\mathrm{d}\mu_T(x)-\int_{\mathbb{R}^d}\varphi(0,x)\,\mathrm{d}\mu_0(x).
\end{aligned}
$$

By construction, μ_t^k converges uniformly to μ_t for $k \to \infty$ so that μ_t indeed satisfies the weak formulation (3.5.3).
We conclude the proof by noticing that the Lipschitz continuity of μ_t^k with respect to time, as well as with respect to initial measure and model functions (see Propositions 3.4.3 and 3.4.6 respectively) is transferred to μ_t via a similar limiting procedure. □

Remark 3.5.6. Note carefully that Theorem 3.5.5 does not guarantee uniqueness of the model solution. In Theorem 4.4.6 we will provide a proof for the missing uniqueness which is based on the theory for generalised models on Polish metric spaces.

3.6 An extension of the linear case- uniform tightness

We conclude this chapter with the study of uniform tightness of solutions to the linear model (3.1.9). In view of Proposition 3.4.3 i) we know that $\|\mu_t\|_{BL^*}$ remains bounded for all $t \in [0, T]$. As $T > 0$ is arbitrary, we can exclude blow-ups in finite time and are thus able to extend the measure solution from $[0, T]$ to $[0, \infty)$.
Our goal is to show that the family $\{\mu_t \mid t \in [0, \infty)\} \subset \mathcal{M}^+(\mathbb{R}^d)$ is tight, i.e. that $\mu_\bullet$ defines a uniformly tight family. We note that for finite and fixed T, tightness follows directly from Proposition 3.4.3 ii) together with the Theorem of Prokhorov 2.1.3. However, for infinite T the mass of the family could still escape towards infinity preventing tightness. To circumvent this, we need to upgrade the regularity of the model functions. In particular, we need uniform L^1 regularity in time, i.e. we assume

Assumptions 3.6.1. The model functions satisfy Assumptions 3.1.1 for all $T > 0$, i.e.

(i) $b \in C_b^0([0, \infty); BL(\mathbb{R}^d; \mathbb{R}^d))$,

(ii) $c \in C_b^0([0, \infty); BL(\mathbb{R}^d))$,

(iii) $\eta \in C_b^0\left([0, \infty); BL(\mathbb{R}^d; \mathcal{M}^+(\mathbb{R}^d))\right)$,

(iv) $N \in C_b^0([0, \infty); \mathcal{M}^+(\mathbb{R}^d))$.

Additionally, we need that

(v)

$$\int_0^\infty \sup_{x\in\mathbb{R}^d} |c(r,x)| \,\mathrm{d}r =: K_c < \infty,$$

(vi)

$$\int_0^\infty \|\eta(r,\cdot)\|_{BL(\mathbb{R}^d;\mathcal{M}^+)} \,\mathrm{d}r =: K_\eta < \infty,$$

(vii)

$$\int_0^\infty \|N(r)\|_{BL^*} \,\mathrm{d}r =: K_N < \infty$$

Remark 3.6.2. Note that in constrast to Assumptions 3.1.1 we explicitly need to assume boundedness in time of the model functions as the time interval is not compact anymore.

Tightness of the measure solutions for all $t > 0$ requires uniform boundedness. As we constructed our measure solution via the dual problem (3.3.2), we first bound $\varphi_{\psi,t}$ uniformly in t by modifying Proposition 3.3.4.

Proposition 3.6.3. *Let $\psi \in BL(\mathbb{R}^d)$ and let the model functions satisfy Assumptions 3.6.1. Then the solution $\varphi_{\psi,t}$ to the dual problem* (3.3.2) *is bounded independently of t by*

$$\|\varphi_{\psi,t}\|_\infty \leq \|\psi\|_\infty \exp(e^{K_c} K_\eta + K_c).$$

Proof. Similar to the proof of Proposition 3.3.4 we first bound the x-component using representation formula (3.3.3). Let $\tau \in [0,t]$. Then

$$
\begin{aligned}
&\sup_{\tilde{x}\in\mathbb{R}^d} |\varphi_{\psi,t}(\tau,\tilde{x})| \\
\leq &\|\psi\|_\infty e^{\int_\tau^t \sup_{x\in\mathbb{R}^d} |c(r,x)|\,\mathrm{d}r} + \int_\tau^t \sup_{\tilde{x}\in\mathbb{R}^d} \int_{\mathbb{R}^d} |\varphi_{\psi,t}(s,y)|\,\mathrm{d}[\eta(s,\tilde{x})](y) e^{\int_\tau^s \sup_{x\in\mathbb{R}^d} |c(r,x)|\,\mathrm{d}r}\,\mathrm{d}s \\
\leq &\|\psi\|_\infty e^{K_c} + e^{K_c} \int_\tau^t \sup_{\tilde{y}\in\mathbb{R}^d} |\varphi_{\psi,t}(s,\tilde{y})| \sup_{\tilde{x}\in\mathbb{R}^d} \int_{\mathbb{R}^d} \mathrm{d}[\eta(s,\tilde{x})](y)\,\mathrm{d}s \\
\leq &\|\psi\|_\infty e^{K_c} + e^{K_c} \int_\tau^t \sup_{\tilde{y}\in\mathbb{R}^d} |\varphi_{\psi,t}(s,\tilde{y})| \|\eta(s,\cdot)\|_{BL(\mathbb{R}^d;\mathcal{M}^+)}\,\mathrm{d}s \\
\leq &\|\psi\|_\infty e^{K_c} + e^{K_c} \int_\tau^t \sup_{\tilde{y}\in\mathbb{R}^d} |\varphi_{\psi,t}(s,\tilde{y})| \|\eta(s,\cdot)\|_{BL(\mathbb{R}^d;\mathcal{M}^+)}\,\mathrm{d}s.
\end{aligned}
$$

Hence, by Gronwall's inequality

$$
\sup_{\tilde{x}\in\mathbb{R}^d} |\varphi_{\psi,t}(\tau,\tilde{x})| \leq e^{K_c}\|\psi\|_\infty e^{\int_\tau^t \|\eta(s,\cdot)\|_{BL(\mathbb{R}^d;\mathcal{M}^+)} e^{K_c}} \leq \|\psi\|_\infty \exp(K_c + e^{K_c} K_\eta),
$$

as desired. □

We are now able to formulate the main result of this section.

Theorem 3.6.4. *Let* $\mu_\bullet : [0,\infty) \to \mathcal{M}^+(\mathbb{R}^d)$ *be the measure solution to* (3.1.9) *for all* $T > 0$. *Under Assumptions 3.6.1 the following statements hold.*

(i) The family $\mathcal{F} := \{\mu_t \mid t \in [0,\infty)\}$ *is uniformly bounded in the flat norm.*

(ii) For all $A \in \mathcal{B}(\mathbb{R}^d)$ *let*

$$
\mu_\infty(A) := \lim_{t\to\infty} \mu_t(A).
$$

Then $\mu_\infty \in \mathcal{M}^+(\mathbb{R}^d)$ *and the closure* $\overline{\mathcal{F}}$ *of* $\mathcal{F}$ *with respect to narrow convergence is given by* $\overline{\mathcal{F}} = \mathcal{F} \cup \{\mu_\infty\}$.

(iii) The family $\mathcal{F}$ *is tight.*

Proof. (i) With the help of Proposition 3.6.3 we can improve the proof of Proposition 3.4.3 or rather of Proposition 3.3.14. In particular, we apply the variational characterisation of the total variation norm (see e.g. [44, F.24]), (3.4.2) and Proposition

3.6.3 which yields

$$\begin{aligned}
\|\mu_t\|_{TV} &= \sup_{\substack{\psi\in C_c^0(\mathbb{R}^d)\\ \|\psi\|_\infty\le 1}} \left(\int_{\mathbb{R}^d} \psi(x)\,\mathrm{d}\mu_t(x) + \int_0^t\int_{\mathbb{R}^d} \varphi_{\psi,t}(\tau,x)\,\mathrm{d}[N(\tau)](x)\,\mathrm{d}\tau\right)\\
&\overset{(*)}{=} \sup_{\substack{\psi\in C_c^1(\mathbb{R}^d)\\ \|\psi\|_\infty\le 1}} \left(\int_{\mathbb{R}^d} \psi(x)\,\mathrm{d}\mu_t(x) + \int_0^t\int_{\mathbb{R}^d} \varphi_{\psi,t}(\tau,x)\,\mathrm{d}[N(\tau)](x)\,\mathrm{d}\tau\right)\\
&\le \sup_{\substack{\psi\in C_c^1(\mathbb{R}^d)\\ \|\psi\|_\infty\le 1}} \int_{\mathbb{R}^d} \varphi_{\psi,t}(0,x)\,\mathrm{d}\mu_0(x) + \sup_{\substack{\psi\in C_c^1(\mathbb{R}^d)\\ \|\psi\|_\infty\le 1}} \|\varphi_{\psi,t}\|_{BL^*}\int_0^\infty \|N(\tau)\|_{BL^*}\,\mathrm{d}\tau\\
&\le \exp(K_c + e^{K_c}K_\eta)\|\mu_0\|_{TV} + \exp(K_c+e^{K_c}K_\eta)K_N\\
&= \exp(e^{K_c}K_\eta + K_c)\left(\|\mu_0\|_{TV} + K_N\right) =: C < \infty.
\end{aligned}$$

In $(*)$ we used that $C_c^1(\mathbb{R}^d)$ is dense in $C_0(\mathbb{R}^d)$ by Stone-Weierstrass Theorem. This proves the uniform boundedness of the family $\mathcal{F}$.
(ii) We first note that (narrow) continuity of the map $t \mapsto \mu_t$ (cf. Proposition 3.4.3) implies independence of μ_∞ from limit sequences $t_n \to \infty$ and μ_∞ is thus well-defined if the limit exists. Clearly the nonnegativity of the μ_t is transfered to μ_∞. However, it is à priori not clear that μ_∞ is σ-additive and hence a measure, as the proof of σ-additivity would involve interchanging two limiting procedures. Instead, this will follow from the Vitali-Hahn-Saks Theorem [18, 4.6.3] if we show that $\lim_{t\to\infty}\mu_t(A)$ exists for all $A \in \mathcal{B}(\mathbb{R}^d)$ and is finite. According to (i) the family $\mathcal{F}$ is uniformly bounded by some constant $C > 0$. We compute for all $A \in \mathcal{B}(\mathbb{R}^d)$

$$\mu_\infty(A) = \lim_{t\to\infty}\mu_t(A) = \lim_{t\to\infty}\int_A 1\,\mathrm{d}\mu_t(x) \le \lim_{t\to\infty}\int_{\mathbb{R}^d} 1\,\mathrm{d}\mu_t(x) \le \lim_{t\to\infty}\|\mu_t\|_{BL^*} \le C.$$

In particular, the limit $\lim_{t\to\infty}\mu_t(A) = \mu_\infty(A)$ exists and is finite. Hence, the Vitali-Hahn-Saks Theorem [18, 4.6.3] implies that μ_∞ indeed defines a measure and is thus in $\mathcal{M}^+(\mathbb{R}^d)$.
Narrow continuity of the map $t \mapsto \mu_t$ implies that $\overline{\mathcal{F}}$ is the closure of $\mathcal{F}$.
(iii) We claim that $\overline{\mathcal{F}}$ is sequentially compact with respect to narrow convergence, so that tightness follows from the Theorem of Prokhorov 2.1.3. Let $(\mu_{t_n})_{n\in\mathbb{N}} \subset \mathcal{F}$ be a sequence. If there is a subsequence $(\tilde{t}_n)_{n\in\mathbb{N}}$ such that $\tilde{t}_n \to \infty$, then by construction $\mu_{\tilde{t}_n} \to \mu_\infty \in \overline{\mathcal{F}}$ and we are done. If no such subsequence exists, then there exists $T > 0$ such that $t_n \in [0,T]$ for all n. By compactness the sequence $(t_n)_{n\in\mathbb{N}} \subset [0,T]$ has a converging subsequence $\left(\tilde{t}_n\right)_{n\in\mathbb{N}}$ with $\tilde{t}_n \to \tilde{t} \in [0,T]$. Then again by narrow continuity $\mu_{\tilde{t}_n} \to \mu_{\tilde{t}} \in \mathcal{F}$ and $\mathcal{F}$ is sequentially compact. □

4 Models on Polish metric spaces

In this chapter, we will use the insights gained from the theory established in Section 3 to develop a precise concept for structured population models on abstract Polish metric spaces (S, d). To this end, we need on the one hand the profound functional analytic knowledge on the spaces $\mathcal{M}$ and $\mathcal{M}^+$ from Section 2. On the other hand, we have to take into account that in general abstract metric spaces do not possess a linear structure. Consequently, we can not rely on the notion of derivatives, in particular not on the method of characteristics which we used to construct the solution of the transport process in (3.3.2). Therefore, we will try to relax the regularity assumptions from Section 3 as much as possible and concentrate on the essential properties and concepts. The ideas presented in this section are based on [46]. We start with generalising the flow of the vector field b.

4.1 Two-parameter families of Lipschitz maps

Definition 4.1.1. *Let (S, d) be a separable metric space. For every pair $(t, \tau) \in \mathbb{R}^2$ we define a **two-parameter family of Lipschitz maps** $X(t, \tau, \cdot) : S \to S$ such that*

(i) the map $t \to X(t, \tau, \cdot)$ is uniformly continuous, i.e.

$$\sup_{\tau \in \mathbb{R}} \sup_{x \in S} d(X(t_1, \tau, x), X(t_2, \tau, x)) \le \omega_X(|t_1 - t_2|)$$

with a modulus of continuity ω_X ($\omega_X : [0, \infty] \to [0, \infty]$ is continuous in 0 and satisfies $\omega_X(0) = 0$).

(ii) $\lim_{t \to \tau} X(t, \tau, \cdot) = \mathrm{Id}$ in $C^0(S)$, i.e. $\lim_{t \to \tau} \sup_{x \in S} d(X(t, \tau, x), x) = 0$,

(iii) $X(t_2, \tau, \cdot) = X(t_2, t_1, \cdot) \circ X(t_1, \tau, \cdot)$ for every $t_1, t_2 \in \mathbb{R}$,

(iv) there exists a locally bounded function $L_X(\cdot)$ with

$$d(X(t, \tau, x_1), X(t, \tau, x_2)) \le L_X(t - \tau)\, d(x_1, x_2).$$

We may also restrict the domain of the time arguments τ and t to some open or closed interval of the form $(0,T)$ or $[0,T]$.

Remark 4.1.2. The two-parameter family of Lipschitz maps captures the for us essential properties of flows generated by vector fields b on $\mathbb{R}^d$ without having to work with an underlying ODE as in (3.2.1). Consequently, the family of Lipschitz maps can be seen as the natural generalisation of flows from $\mathbb{R}^d$ to generic metric spaces (S,d). In contrast to [46], we only assume Lipschitz continuity of X and not that it is actually a bi-Lipschitz homeomorphism. The latter assumption was strongly inspired by the properties of the flow X_b but proved to be far more than sufficient.

Remark 4.1.3. In accordance with our notation for flows of vector fields, we interpret t and τ in $X(t,\tau,x)$ as **current time** and as **initial time point**, respectively, while x represents **initial position**. Note that we did not assume $\tau \leq t$, so that we can represent backward flows using $t < \tau$.

Our first lemma shows that the flow of a vector field is indeed an example of a two-parameter family of Lipschitz maps. The result follows from the more general Lemma 4.4.2 for the nonlinear case and thus we omit the proof here.

Lemma 4.1.4. *Let $b \in L^1((0,T);BL(\mathbb{R}^d))$, i.e. $\int_0^T \|b(t,\cdot)\|_{BL}\,\mathrm{d}t < \infty$. Then the unique solution $X_b(t;\tau,x)$ of*

$$\partial_t X_b(t;\tau,x) = b(t, X_b(t;\tau,x)), \qquad X_b(\tau;\tau,x) = x$$

defines a two-parameter family of Lipschitz maps on $\mathbb{R}^d$.

4.2 The generalised linear structured population model

Now that we introduced a suitable generalisation of the flow, our next step is to formulate and investigate a corresponding version of the population model (3.5.1) for Polish metric spaces (S,d) as underlying state space. Due to the lack of a vector space structure, it is not possible to formulate an explicit PDE model on (S,d). We thus have to find a different approach that works without differential equations.

Analogously to Section 3 we first treat the simpler linear case and introduce the following model functions

$$\begin{aligned} &c : [0,T] \times \mathbb{R}^d \times \mathcal{M}^+(\mathbb{R}^d) \to \mathbb{R}, \qquad && \eta : [0,T] \times S \to \mathcal{M}^+(S), \\ &N : [0,T] \to \mathcal{M}^+(S), && X : [0,T] \times [0,T] \times S \to S. \end{aligned}$$

Without a PDE model to rely on, the model functions first have to stand alone without a governing equation that gives them meaning. However, as they are clearly inspired by the models introduced in Section 3, the interpretation of c, η and N is similar to their counterparts in the Euclidean setting. In particular, c represents a **growth term**, η is a **spread of heterogeneity** and N is a **state-independent influx**. Function X is the **generalisation of the flow of the vector field** b in the sense of Definition 4.1.1. Our model functions should satisfy the following:

Assumptions 4.2.1 (Linear Model)**.** The model functions c, η, N and X satisfy

(i) $c \in L^1((0,T); BL(S))$,

(ii) $\eta \in L^1\left((0,T); BL(S; \mathcal{M}^+(S))\right)$,

(iii) $N \in L^1\left((0,T); \mathcal{M}^+(S)\right)$,

(iv) X is a two-parameter family of Lipschitz maps on (S,d).

Note that in comparison to Assumption 3.1.1 the abstract setting requires less regularity in time of the model functions. Similar to Notation 3.1.2 we shortly explain the assumptions.

Notation 4.2.2. *Again we restrict ourselves to the most complicated model function η. For fixed $t \in (0,T)$ and $x \in S$, $\eta(t,x)$ is a measure in $\mathcal{M}^+(S)$, and for fixed t the map $x \mapsto \eta(t,x) \in \mathcal{M}^+(S)$ is bounded Lipschitz, i.e.*

$$\|\eta(t,\cdot)\|_{BL(S;\mathcal{M}^+)} := \max\left\{ \sup_{x\in S} \|\eta(t,x)\|_{BL^*}, |\eta(t,\cdot)|_{\mathbf{Lip}} \right\} < \infty$$

with Lipschitz constant given by

$$|\eta(t,\cdot)|_{\mathbf{Lip}} = \sup_{x_1 \neq x_2} \frac{\rho_F(\eta(t,x_1), \eta(t,x_2))}{d(x_1,x_2)}.$$

Furthermore, $\|\eta(t,\cdot)\|_{BL(S;\mathcal{M}^+)}$ *is integrable in time, i.e.*

$$\|\eta\|_{L^1_T(BL(S;\mathcal{M}^+))} := \int_0^T \|\eta(t,\cdot)\|_{BL(S;\mathcal{M}^+)}\,\mathrm{d}t < \infty.$$

Consequently, for a.e. $t \in (0,T)$ *the expression* $\|\eta(t,\cdot)\|_{BL(S;\mathcal{M}^+)}$ *is also finite.*

In the Euclidean case $S = \mathbb{R}^d$ we can directly use the flow generated by the vector field b instead of considering the two-parameter family of Lipschitz maps. In Assumption 4.4.1 we provide assumptions on b in the nonlinear case which are similar but slightly weaker than Assumptions 3.5.1.
As a next step, we introduce the generalised version of the linear structured population model. The basic idea is that we omit the model formulation completely and simply state a suitable notion of solution to our generalised problem which is motivated by the implicit integral representation (3.3.3). See Remark 4.4.7 for a discussion.

Definition 4.2.3 (Generalised Model)**.** *Let* (S,d) *be Polish metric space. A family of measures* $\mu_\bullet := \{\mu_t\}_{t\in[0,T]} \subset \mathcal{M}^+(S)$ *is called a* ***(generalised) solution to the linear structured population model*** *on* (S,d) *with initial measure* $\mu_0 \in \mathcal{M}^+(S)$ *and model functions* (c,η,N,X) *satisfying Assumptions 4.2.1, if* $t \mapsto \mu_t$ *is narrowly continuous and* μ_t *satisfies*

$$\begin{aligned}\mu_t &= X(t,0,\cdot)_\#\left(\mu_0(\cdot)e^{\int_0^t c(s,X(s,0,\cdot))\,\mathrm{d}s}\right)\\ &+ \int_0^t X(t,\tau,\cdot)_\#\left(\int_S [\eta(\tau,y)(\cdot)]\,\mathrm{d}\mu_\tau(y)e^{\int_\tau^t c(s,X(s,\tau,\cdot))\,\mathrm{d}s}\right)\mathrm{d}\tau \\ &+ \int_0^t X(t,\tau,\cdot)_\#\left(N(\tau)(\cdot)e^{\int_\tau^t c(s,X(s,\tau,\cdot))\,\mathrm{d}s}\right)\mathrm{d}\tau.\end{aligned} \tag{4.2.1}$$

The integrals appearing in the second and third term of formula (4.2.1) should be understood in the Bochner sense. They are well-defined according to the following proposition.

Proposition 4.2.4 (Rigorous definition of integrals)**.** *Under Assumptions 4.2.1 the inner and outer Bochner integrals appearing in* (4.2.1) *are well-defined measures in* $\mathcal{M}^+(S)$.

Proof. We need to ensure that the conditions for all appearing functions to be Bochner integrable are met. In order to avoid the associated technical subtleties such as strong measurability, we assume that the target spaces of our functions are separable Banach spaces. In this case Pettis Theorem [164, Theorem, Section V, §4] guarantees that the concept of strong measurability coincides with the easier to handle weak measurability. Now according to the Theorems 2.3.1 and 2.3.8 from Section 2 $\mathcal{M}^+(S)$ is a Polish metric space, but not a vector space and as such not Banach. Hence, instead we will work in the space

$$E := \overline{\mathcal{M}(S)}^{\|\cdot\|_{BL(S)^*}},$$

i.e. the closure of $\mathcal{M}(S)$ as a subspace of $BL(S)^*$. By construction, E is a separable Banach space as the closure of a separable space and as a closed subset of the Banach space $BL(S)^*$. As already pointed out above, the concepts of weak and strong measurability hence coincide on E. In Theorem 2.3.2 we characterised its dual space to be $E^* = BL(S)$.
We start with the inner integral of the second term in (4.2.1) and have to check that the map $S \ni x \mapsto \eta(t,x) \in E$ is weakly measurable, i.e. that for any $f \in E^*$ and a.e. $t \in (0,T)$, the map

$$S \ni x \mapsto \langle f, \eta(t,x)\rangle_{E^*,E}$$

is measurable where $\langle \cdot,\cdot\rangle_{E^*,E}$ denotes the usual dual pairing. As η is assumed to be in $BL(S;\mathcal{M}^+(S))$, this map is not only measurable but even Lipschitz continuous since for all $x_1, x_2 \in S$

$$\begin{aligned}\left|\big\langle f, \eta(t,x_1) - \eta(t,x_2)\big\rangle_{E^*,E}\right| \le& \|f\|_{E^*}\, \|\eta(t,x_1) - \eta(t,x_2)\|_{BL(S)^*}\\ \le& \|f\|_{E^*} |\eta(t,\cdot)|_{\mathbf{Lip}}\, d(x_1,x_2).\end{aligned}$$

Hence, we conclude that $S \ni x \mapsto \eta(t,x)$ is strongly measurable. Additionally, we see that

$$\int_S \|\eta(t,x)\|_{BL^*}\, \mathrm{d}\mu_t(x) \le \|\eta(t,\cdot)\|_{BL(S;\mathcal{M}^+)} \|\mu_t\|_{BL^*} < \infty \tag{4.2.2}$$

which is finite as the narrow continuity of $[0,T] \ni \tau \mapsto \mu_\tau$ implies $\sup_{\tau\in[0,T]} \|\mu_\tau\|_{BL^*} < \infty$. So according to Theorem 1 in Section V §5 in Ref. [164] the map $x \mapsto \eta(t,x)$ is μ_t-integrable for a.e. $t \in (0,T)$ and for a.e. $\tau \in (0,T)$, $I(\tau) = \int_S [\eta(\tau,y)(\cdot)]\, \mathrm{d}\mu_\tau(y)$

is a well defined Bochner integral on E. From (4.2.2) we also conclude that

$$\|I(t)\|_{BL^*} = \left\| \int_S \eta(t,x)\,\mathrm{d}\mu_t(x) \right\|_{BL^*} \leq \|\eta(t,\cdot)\|_{BL(S;\mathcal{M}^+)} \|\mu_t\|_{BL^*}.$$

So far, we have proven that $I(\tau) \in E$. Nevertheless, for (4.2.1) to make sense, we need $I(\tau) \in \mathcal{M}^+(S)$. Thus, we invoke Theorem 2.3.9 to show that $I(\tau)$ is actually a nonnegative measure. To this end, we first show $I(\tau) \in BL(S)^*_+$. Let $f \in BL(S)$ with $f \geq 0$. As Bochner integrals commute with bounded, linear operators (see [164, V §5, Corollary 2]), we compute for any $\nu \in \mathcal{M}^+(S)$

$$\left\langle \int_S \eta(t,x)(\cdot)\,\mathrm{d}\nu(x), f \right\rangle_{BL^*,BL} = \int_S \langle \eta(t,x)(\cdot), f \rangle_{BL^*,BL}\,\mathrm{d}\nu(y) \geq 0,$$

since $\eta(t,x) \in \mathcal{M}^+(S)$. So $I(\tau) \in E \cap BL(S)^*_+$ which happens to be $\mathcal{M}^+(S)$ by Theorem 2.3.9. This concludes the proof for the inner integral of the second term in (4.2.1). However, we can proceed analogously with the outer integral

$$J = \int_0^t X(t,\tau,\cdot)_\# \left[I(\tau)(\cdot) e^{\int_\tau^t c(s,\,X(s,\tau,\cdot))\,\mathrm{d}s} \right] \mathrm{d}\tau.$$

In particular, the map $\tau \mapsto X(t,\tau,\cdot)_\# \left[I(\tau)(\cdot) e^{\int_\tau^t c(s,\,X(s,\tau,\cdot))\,\mathrm{d}s} \right]$ is weakly measurable and the map

$$\tau \mapsto \left\langle f, X(t,\tau,\cdot)_\# \left[I(\tau)(\cdot) e^{\int_\tau^t c(s,\,X(s,\tau,\cdot))\,\mathrm{d}s} \right] \right\rangle_{E^*,E}$$

is continuous. A similar argument as for $I(\tau)$ shows that J is also a nonnegative Radon measure on S. The integrals appearing in the third term in (4.2.1) can be analysed analogously. $\square$

Remark 4.2.5. Consider the conservative case when $c, \eta, N \equiv 0$, i.e. the initial measure μ_0 is just moved according to the map $X(t,\cdot,0)$. In this case the solution μ_t satisfies $\mu_t = X(t,\cdot,0)_\# \mu_0$ which is exactly the solution to the optimal transport problem with transport map $X(t,\cdot,0)$ (see [159, p.16-17]). So from an optimal transport perspective formula (4.2.1) provides a substantial generalisation to the simple optimal transport case.

The implicit formulation for μ_t appearing in Definition 4.2.3 is quite impractical to explicitly calculate integrals with it. So before we treat the well-posedness theory for our generalised model, we first show how to compute integrals of the form

$\int_S \psi(x)\,\mathrm{d}\mu_t(x)$ for measures μ_t given by the implicit equation (4.2.1). To this end, we decouple the left and the right hand side of (4.2.1).

Lemma 4.2.6. *Consider a narrowly continuous map $\nu_\bullet : [0,T] \to \mathcal{M}^+(S)$ and let $\mu_0 \in \mathcal{M}^+(S)$. For $t \in [0,T]$ we define a family of measures μ_t by*

$$\begin{aligned}
\mu_t &:= X(t,0,\cdot)_\# \left(\mu_0(\cdot) e^{\int_0^t c(s,\, X(s,0,\cdot))\,\mathrm{d}s}\right) \\
&+ \int_0^t X(t,\tau,\cdot)_\# \left(\int_S [\eta(\tau,y)(\cdot)]\,\mathrm{d}\nu_\tau(y) e^{\int_\tau^t c(s,\, X(s,\tau,\cdot))\,\mathrm{d}s}\right) \mathrm{d}\tau \\
&+ \int_0^t X(t,\tau,\cdot)_\# \left(N(\tau)(\cdot) e^{\int_\tau^t c(s,\, X(s,\tau,\cdot))\,\mathrm{d}s}\right) \mathrm{d}\tau.
\end{aligned} \tag{4.2.3}$$

Then for any $\psi \in BL(S)$ the integral with respect to μ_t can be computed as follows

$$\begin{aligned}
\int_S \psi(x)\,\mathrm{d}\mu_t(x) &= \int_S \psi(X(t,0,x)) e^{\int_0^t c(s,\, X(s,0,x))\,\mathrm{d}s}\,\mathrm{d}\mu_0(x) \\
&+ \int_0^t \int_S \int_S \psi(X(t,\tau,y)) e^{\int_\tau^t c(s,\, X(s,\tau,y))\,\mathrm{d}s}\,\mathrm{d}[\eta(\tau,x)](y)\,\mathrm{d}\nu_\tau(x)\,\mathrm{d}\tau \\
&+ \int_0^t \int_S \psi(X(t,\tau,x)) e^{\int_\tau^t c(s,\, X(s,\tau,x))\,\mathrm{d}s}\,\mathrm{d}[N(\tau)](x)\,\mathrm{d}\tau.
\end{aligned} \tag{4.2.4}$$

Proof. We treat each term of μ_t individually and start with the first one. A direct application of the change of variable formula (2.1.1) for push-forward measures to the first term in (4.2.3) leads to the first term on the right-hand side of (4.2.4).
Before we continue, we note that according to Theorem 2.3.2 $\psi \in BL(S)$ implies $\psi \in E^*$. Hence, as linear operations commute with the Bochner integral (see [164, V §5, Corollary 2]), we compute

$$\begin{aligned}
&\int_S \psi(y)\,\mathrm{d}\left[\int_0^t X(t,\tau,\cdot)_\# \left(\int_S [\eta(\tau,x)(\cdot)]\,\mathrm{d}\nu_\tau(x) e^{\int_\tau^t c(s,\, X(s,\tau,\cdot))\,\mathrm{d}s}\right) \mathrm{d}\tau\right](y) \\
&= \int_0^t \int_S \psi(y)\,\mathrm{d}\left[X(t,\tau,\cdot)_\# \left(\int_S [\eta(\tau,x)(\cdot)]\,\mathrm{d}\nu_\tau(x) e^{\int_\tau^t c(s,\, X(s,\tau,\cdot))\,\mathrm{d}s}\right)\right](y)\,\mathrm{d}\tau \\
&= \int_0^t \int_S \psi(X(t,\tau,y))\,\mathrm{d}\left[\int_S [\eta(\tau,x)(\cdot)]\,\mathrm{d}\nu_\tau(x) e^{\int_\tau^t c(s,\, X(s,\tau,\cdot))\,\mathrm{d}s}\right](y)\,\mathrm{d}\tau \\
&= \int_0^t \int_S \int_S \psi(X(t,\tau,y)) e^{\int_\tau^t c(s,\, X(s,\tau,y))\,\mathrm{d}s}\,\mathrm{d}[\eta(\tau,x)]\,(y)\,\mathrm{d}\nu_\tau(x)\,\mathrm{d}\tau.
\end{aligned}$$

The third term can be handled in a similar manner. □

Analogously to Lemma 3.3.6 in the Euclidean case, we need some auxiliary regularity and boundedness results to streamline the upcoming computations. The ideas of the proofs are sufficiently close, so that we omit the calculations here and instead refer to Lemma 3.17 in Ref. [44].

Lemma 4.2.7. *Suppose functions X and c satisfy Assumption 4.2.1. For all test functions ψ with $\|\psi\|_{BL} \leq 1$.*

(i) The map $[\tau, T] \ni t \mapsto G^{\psi,X,c}_{\tau,t}(x) := \psi(X(t,\tau,x))e^{\int_\tau^t c(s,X(s,\tau,x))\,\mathrm{d}s}$ is uniformly continuous, i.e. there exists a modulus of continuity $\omega_{X,c}$ such that

$$\left|G^{\psi,X,c}_{\tau,t_2}(x) - G^{\psi,X,c}_{\tau,t_1}(x)\right| \leq \omega_{X,c}\left(|t_2 - t_1|\right). \tag{4.2.5}$$

(ii) For any $0 \leq \tau \leq t \leq T$, the map $S \ni x \mapsto G^{\psi,X,c}_{\tau,t}(x)$ is in BL(S) and $\|G^{\psi,X,c}_{\tau,t}\|_{BL}$ can be bounded independently of ψ by

$$C^{X,c}_{\tau,t} := e^{\int_0^T \|c(s,\cdot)\|_\infty\,\mathrm{d}s}\left[L_X(t-\tau) + \int_\tau^t |c(s,\cdot)|_{\mathbf{Lip}}\, L_X(s-\tau)\,\mathrm{d}s\right], \tag{4.2.6}$$

where he function L_X has been introduced in Definition 4.1.1.

We will establish existence and uniqueness of the generalised solutions with the help of Banach Fixed Point Theorem applied to an operator built on (4.2.1). Due to the low L^1 regularity in time of the model functions, the standard Bielecki metric is insufficient for proving contractivity in this case. Thus, based on the idea introduced in Ref. [97] we define a new complete metric on $C^0([0,T];\mathcal{M}^+(S))$ as follows: For an a.e. nonnegative $f \in L^1(0,T)$ and $\lambda > 0$ we introduce

$$\rho_{\lambda,f}(\mu_\bullet, \nu_\bullet) = \sup_{t\in[0,T]} \left[e^{-\lambda \int_0^t f(u)\,\mathrm{d}u} \rho_F(\mu_t, \nu_t)\right]. \tag{4.2.7}$$

Choosing $f = 1$ reduces (4.2.7) to a metric related to the standard Bielecki norm, see (3.3.50).
Theorem 2.3.8 states that the space $(\mathcal{M}^+(S), \rho_F)$ is complete and consequently also $C^0([0,T];\mathcal{M}^+(S))$ equipped with the natural metric

$$\rho_\infty(\mu_\bullet, \nu_\bullet) := \sup_{t\in[0,T]} \rho_F(\mu_t, \nu_t). \tag{4.2.8}$$

The equivalence of $\rho_{\lambda,f}$ to the metric ρ_∞ directly implies that the space

$$(C^0([0,T];\mathcal{M}^+(S)), \rho_{\lambda,f})$$

is complete as well. We remark that the completeness of $(\mathcal{M}^+(S), \rho_F)$ is crucial here: As we have mentioned in Section 2.3 the space of signed measures $\mathcal{M}(S)$ is only complete under the flat metric if S is uniformly discrete, so that in general we can not work with $C^0([0,T];\mathcal{M}(S))$.
Now that we have a suitable complete metric space, we construct our generalised solution as a fixed point of the following operator

$$F : C^0([0,T];\mathcal{M}^+(S)) \to C^0([0,T];\mathcal{M}^+(S)).$$

For a measure-valued map $\mu_\bullet \in C^0([0,T];\mathcal{M}^+(S))$, define $(F\mu_\bullet)_\bullet$ by the right-hand side of (4.2.1), i.e.

$$\begin{aligned}
&(F\mu_\bullet)_t = X(t,0,\cdot)_\#(\mu_0(\cdot)e^{\int_0^t c(s,\,X(s,0,\cdot))\,\mathrm{d}s}) \\
&+ \int_0^t X(t,\tau,\cdot)_\# \left(\int_S [\eta(\tau,y)(\cdot)]\,\mathrm{d}\mu_\tau(y) e^{\int_\tau^t c(s,\,X(s,\tau,\cdot))\,\mathrm{d}s}\right)\mathrm{d}\tau \\
&+ \int_0^t X(t,\tau,\cdot)_\# \left(N(\tau)(\cdot) e^{\int_\tau^t c(s,\,X(s,\tau,\cdot))\,\mathrm{d}s}\right)\mathrm{d}\tau.
\end{aligned} \tag{4.2.9}$$

First, we will show that that $(F\mu_\bullet)_\bullet : [0,T] \to \mathcal{M}^+(S)$ is continuous.

Lemma 4.2.8. *Let $\mu_\bullet \in C^0([0,T];\mathcal{M}^+(S))$ and let $(F\mu_\bullet)_\bullet$ be defined via* (4.2.9)*. Then, $t \mapsto (F\mu_\bullet)_t$ is continuous as well, i.e. $(F\mu_\bullet)_\bullet \in C^0([0,T];\mathcal{M}^+(S))$.*

Proof. Let $\mu_\bullet \in C^0([0,T];\mathcal{M}^+(S))$ and $\psi \in BL(S)$ with $\|\psi\|_{BL} \le 1$. Moreover, for $0 \le t_1 < t_2 \le T$ let $G^{\psi,X,c}_{\tau,t}$ be as in Lemma 4.2.7. Then we compute with (4.2.4) rewritten in terms of $G^{\psi,X,c}_{\tau,t}$

$$\begin{aligned}
\int_S \psi(x)\,\mathrm{d}[(F\mu)_{t_2} - (F\mu)_{t_1}](x) = &\int_S \left[G^{\psi,X,c}_{0,t_2}(x) - G^{\psi,X,c}_{0,t_1}(x)\right]\mathrm{d}\mu_0(x) \\
&+ \int_0^{t_2}\int_S\int_S G^{\psi,X,c}_{\tau,t_2}(y)\,\mathrm{d}[\eta(\tau,x)](y)\,\mathrm{d}\mu_\tau(x)\,\mathrm{d}\tau \\
&- \int_0^{t_1}\int_S\int_S G^{\psi,X,c}_{\tau,t_1}(y)\,\mathrm{d}[\eta(\tau,x)](y)\,\mathrm{d}\mu_\tau(x)\,\mathrm{d}\tau \\
&+ \int_0^{t_2}\int_S\int_S G^{\psi,X,c}_{\tau,t_2}(x)\,\mathrm{d}[N(\tau)](x)\,\mathrm{d}\tau - \int_0^{t_1}\int_S\int_S G^{\psi,X,c}_{\tau,t_1}(x)\,\mathrm{d}[N(\tau)](x)\,\mathrm{d}\tau \\
:= &A + B_1 - C_1 + B_2 - C_2.
\end{aligned} \tag{4.2.10}$$

Due to estimate (4.2.5),

$$|A| \leq \omega_{X,c}\left(|t_2 - t_1|\right) \|\mu_0\|_{BL^*},$$

so that $A \to 0$ as $t_2 \to t_1$, independently of ψ with $\|\psi\|_{BL} \leq 1$.
Next, we rewrite the term $B_1 - C_1$ as

$$\begin{aligned} B_1 - C_1 = &\int_{t_1}^{t_2} \int_S \int_S G^{\psi,X,c}_{\tau,t_2}(y)\, \mathrm{d}[\eta(\tau,x)](y)\, \mathrm{d}\mu_\tau(x)\, \mathrm{d}\tau \\ &+ \int_0^{t_1} \int_S \int_S \left[G^{\psi,X,c}_{\tau,t_2}(y) - G^{\psi,X,c}_{\tau,t_1}(y)\right] \mathrm{d}[\eta(\tau,x)](y)\, \mathrm{d}\mu_\tau(x)\, \mathrm{d}\tau := E_1 + D_1. \end{aligned}$$

The term E_1 can be estimated by bound (4.2.6)

$$\begin{aligned} E_1 &\leq \int_{t_1}^{t_2} \int_S C^{X,c}_{\tau,t_2} \|\eta(\tau,x)\|_{BL^*}\, \mathrm{d}\mu_\tau(x)\, \mathrm{d}\tau \leq \int_{t_1}^{t_2} C^{X,c}_{\tau,t_2} \left[\sup_{x\in S} \|\eta(\tau,x)\|_{BL^*}\right] \|\mu_\tau\|_{BL^*}\, \mathrm{d}\tau \\ &\leq \left(\int_{t_1}^{t_2} \|\eta(\tau,\cdot)\|_{BL(BL^*)}\, \mathrm{d}\tau\right) \sup_{\tau\in[t_1,\,t_2]} C^{X,c}_{\tau,t_2} \|\mu_\tau\|_{BL^*}. \end{aligned}$$

By Assumptions 4.2.1 or rather Definition 4.1.1 L_X is locally bounded so that $\sup_{\tau\in[t_1,\,t_2]} C^{X,c}_{\tau,t_2}$ is finite. Furthermore, the map $\tau \mapsto \|\eta(\tau,\cdot)\|_{BL(BL^*)}$ is in $L^1(0,T)$ and $\sup_{\tau\in[t_1,t_2]} \|\mu_\tau\|_{BL^*}$ is finite by the continuity assumption on $\mu_\bullet$. From this we conclude with Dominated Convergence Theorem that $E_1 \to 0$ as $t_2 \to t_1$ independently of ψ with $\|\psi\|_{BL} \leq 1$. Finally, using (4.2.5)

$$\begin{aligned} |D_1| \leq &\omega_{X,c}\left(|t_2 - t_1|\right) \int_0^{t_1} \int_S \|\eta(\tau,x)\|_{BL^*}\, \mathrm{d}\mu_\tau(x)\, \mathrm{d}\tau \\ \leq &\omega_{X,c}\left(|t_2 - t_1|\right) \|\eta\|_{L^1_T(BL(BL^*))} \sup_{\tau\in[0,t_1]} \|\mu_\tau\|_{BL^*}, \end{aligned}$$

implying that also $F_1 \to 0$ as $t_2 \to t_1$, independently of ψ with $\|\psi\|_{BL} \leq 1$. The same procedure can be applied to the term $(B_2 - C_2)$. As the above computations are independent of $\psi \in BL(S)$ with $\|\psi\|_{BL^*} \leq 1$ equation (4.2.10) implies that

$$\lim_{t_2\to t_1} \rho_F\left((F\mu)_{t_2}, (F\mu)_{t_1}\right) \to 0,$$

showing that indeed $(F\mu_\bullet)_\bullet \in C^0([0,T]; \mathcal{M}^+(S))$. □

We are ready able to show existence and uniqueness of generalised solutions for the linear model (4.2.1).

Theorem 4.2.9. *Under Assumption 4.2.1, there exists a unique solution to the linear problem* (4.2.1) *with initial measure* $\mu_0 \in \mathcal{M}^+(S)$.

Proof. Our goal is to show that F defined by (4.2.9) is a contraction, so that existence and uniqueness follow directly from Banach Fixed Point Theorem.
Thus, consider two measure-valued maps $\mu_\bullet^{(1)}, \mu_\bullet^{(2)} \in C^0([0,T]; \mathcal{M}^+(S))$ with initial measure

$$\mu_0^{(1)} = \mu_0^{(2)} = \mu_0 \in \mathcal{M}^+(S).$$

We first note that Assumption 4.2.1 implies for all $\psi \in BL(S)$ and all $t \in [0,T]$ that the map

$$S \ni x \mapsto \int_S \psi(y) \,\mathrm{d}[\eta(t,x)](y) \tag{4.2.11}$$

is in $BL(S)$. The corresponding $\|\cdot\|_{BL}$ norm can be bounded by $\|\psi\|_{BL} \, \|\eta(t,\cdot)\|_{BL(S;\mathcal{M}^+)}$. Thus, we compute using (4.2.4) and the bound (4.2.6)

$$\begin{aligned}
&\rho_F\left((F\mu_\bullet^{(1)})_t, (F\mu_\bullet^{(2)})_t\right) \\
&\le \sup_{\|\psi\|_{BL}\le 1} \int_0^t \int_S \int_S \psi(X(t,\tau,y))\, e^{\int_\tau^t c(s,X(s,\tau,y))\,\mathrm{d}s} \,\mathrm{d}[\eta(\tau,x)](y) \,\mathrm{d}\left[\mu_\tau^{(1)} - \mu_\tau^{(2)}\right](x) \,\mathrm{d}\tau \\
&\le \int_0^t 2\, C_{\tau,t}^{X,c} \, \|\eta(\tau,\cdot)\|_{BL(S;\mathcal{M}^+)} \, \rho_F\left(\mu_\tau^{(1)}, \mu_\tau^{(2)}\right) \mathrm{d}\tau.
\end{aligned}$$

Let $C = \sup_{0\le t\le T} \sup_{0\le\tau\le t} C_{\tau,t}^{X,c}$ which is finite as L_X is locally bounded by Assumptions 4.2.1. Then

$$\rho_F((F\mu_\bullet^{(1)})_t, (F\mu_\bullet^{(2)})_t) \le C \int_0^t \|\eta(\tau,\cdot)\|_{BL(S;\mathcal{M}^+)} \, \rho_F\left(\mu_\tau^{(1)}, \mu_\tau^{(2)}\right) \mathrm{d}\tau.$$

In the next step, we apply the generalised Bielecki norm $\rho_{\lambda,f}$ introduced in (4.2.7), where suitable $\lambda > 0$ and $f \in L^1(0,T)$ are chosen later. We compute

$$\rho_{\lambda,f}(F\mu_\bullet^{(1)}, F\mu_\bullet^{(2)}) = \sup_{t\in[0,T]} \left[e^{-\lambda\int_0^t f(u)\,\mathrm{d}u} \rho_F((F\mu^{(1)})_t, (F\mu^{(2)})_t)\right]$$
$$\leq C \sup_{t\in[0,T]} e^{-\lambda\int_0^t f(u)\,\mathrm{d}u} \int_0^t \|\eta(\tau,\cdot)\|_{BL(S;\mathcal{M}^+)}\ \rho_F\left(\mu_\tau^{(1)}, \mu_\tau^{(2)}\right) \mathrm{d}\tau$$
$$\leq C\rho_{\lambda,f}(\mu_\bullet^{(1)}, \mu_\bullet^{(2)}) \sup_{t\in[0,T]} e^{-\lambda\int_0^t f(u)\,\mathrm{d}u} \int_0^t \|\eta(\tau,\cdot)\|_{BL(S;\mathcal{M}^+)}\ e^{\lambda\int_0^\tau f(u)\,\mathrm{d}u}\,\mathrm{d}\tau.$$

We remark that the integral can be simplified by a clever choice of f. Indeed, for $f(\tau) = \|\eta(\tau,\cdot)\|_{BL(S;\mathcal{M}^+)} \in L^1((0,T))$ we note that the map $\tau \mapsto \lambda f(\tau) e^{\lambda\int_0^\tau f(u)\,\mathrm{d}u}$ is the derivative of $\tau \mapsto e^{\lambda\int_0^\tau f(u)\,\mathrm{d}u}$ due to Lebesgue Differentiation Theorem (see Theorem 6 in Appendix E in Ref. [52]), so that

$$\rho_{\lambda,f}(F\mu_\bullet^{(1)}, F\mu_\bullet^{(2)}) \leq \frac{C}{\lambda}\rho_{\lambda,f}(\mu_\bullet^{(1)}, \mu_\bullet^{(2)}) \sup_{t\in[0,T]} e^{-\lambda\int_0^t f(u)\,\mathrm{d}u}\left[e^{\lambda\int_0^t f(u)\,\mathrm{d}u} - 1\right]$$
$$= \frac{C}{\lambda}\rho_{\lambda,f}(\mu_\bullet^{(1)}, \mu_\bullet^{(2)}) \sup_{t\in[0,T]}\left[1 - e^{-\lambda\int_0^t f(u)\,\mathrm{d}u}\right] = \frac{C}{\lambda}\rho_{\lambda,f}(\mu_\bullet^{(1)}, \mu_\bullet^{(2)})\left[1 - e^{-\lambda\int_0^T f(u)\,\mathrm{d}u}\right]$$
$$\leq \frac{C}{\lambda}\rho_{\lambda,f}(\mu_\bullet^{(1)}, \mu_\bullet^{(2)}).$$

Setting $\lambda = 2\,C$ finishes the proof. □

Next, we prove an *à* priori bound on the generalised solution before we show the continuous dependency of solutions on initial conditions.

Lemma 4.2.10. *Suppose* $[0,T] \ni t \mapsto \mu_t \in \mathcal{M}(S)$ *is the generalised solution to* (4.2.1). *Then* $\|\mu_t\|_{BL^*}$ *is uniformly bounded for all* $t \in [0,T]$.

Proof. We plug $\psi = 1$ into formula (4.2.4) which yields

$$\|\mu_t\|_{BL^*} \leq \|\mu_0\|_{BL^*}\, e^{\int_0^t \|c(s,\cdot)\|_\infty\,\mathrm{d}s}$$
$$+ \int_0^t \left(\sup_{x\in S}\|\eta(\tau,x)\|_{BL^*}\|\mu_\tau\|_{BL^*} + \|N(\tau)\|_{BL^*}\right) e^{\int_\tau^t \|c(s,\cdot)\|_\infty\,\mathrm{d}s}\,\mathrm{d}\tau$$

so that the claim follows by Gronwall's inequality. □

Proposition 4.2.11. *Let* $\mu_\bullet$, $\nu_\bullet$ *be generalised solutions to* (4.2.1) *with initial measures* μ_0 *and* ν_0 *respectively. Then there exists a constant* C *depending on* X *and* c *such that*

$$\rho_F(\mu_t, \nu_t) \leq C\rho_F(\mu_0, \nu_0)\, e^{C\int_0^t \|\eta(\tau,\cdot)\|_{BL(S;\mathcal{M}^+)}\,\mathrm{d}\tau},$$

i.e. the solutions are continuous with respect to initial conditions.

Proof. We apply identity (4.2.4), the bound (4.2.6) and use the boundedness of (4.2.11) to see

$$\begin{aligned}
\rho_F(\mu_t,\nu_t) &\leq \sup_{\|\psi\|_{BL}\leq 1}\int_S \psi(X(t,0,x))e^{\int_0^t c(s,\,X(s,0,x))\,\mathrm{d}s}\,\mathrm{d}[\mu_0-\nu_0]\,(x)\\
&\quad+\sup_{\|\psi\|_{BL}\leq 1}\int_0^t\int_S\int_S \psi(X(t,\tau,y))e^{\int_\tau^t c(s,\,X(s,\tau,y))\,\mathrm{d}s}\,\mathrm{d}[\eta(\tau,x)](y)\,\mathrm{d}[\mu_\tau-\nu_\tau]\,(x)\,\mathrm{d}\tau\\
&\leq C_{0,t}^{X,c}\,\rho_F(\mu_0,\nu_0)+\int_0^t C_{\tau,t}^{X,c}\,\|\eta(\tau,\cdot)\|_{BL(S;\mathcal{M}^+)}\;\rho_F(\mu_\tau,\nu_\tau)\,\mathrm{d}\tau\\
&\leq C\,\rho_F(\mu_0,\nu_0)+C\int_0^t \|\eta(\tau,\cdot)\|_{BL(S;\mathcal{M}^+)}\;\rho_F(\mu_\tau,\nu_\tau)\,\mathrm{d}\tau,
\end{aligned}$$

where $C=\sup_{0\leq t\leq T}\sup_{0\leq\tau\leq t}C_{\tau,t}^{X,c}$ which is finite as L_X is locally bounded by Assumptions 4.2.1. We thus conclude the proof with an application of Gronwall's inequality. $\square$

With similar computations (and many triangle inequalites) we can prove that the generalised solutions also depend continuously on the model functions. We omit the proof here and simply state the result. The lengthy computations can be found in [44, Lemma 3.26].

Proposition 4.2.12. *Let* $\mu^{c,\eta,N,X}$ *and* $\mu^{\tilde{c},\tilde{\eta},\tilde{N},\tilde{X}}$ *be the two generalised solutions of* (4.2.1) *with initial condition* $\mu_0^{c,\eta,N,X}=\mu_0^{\tilde{c},\tilde{\eta},\tilde{N},\tilde{X}}=\mu_0\in\mathcal{M}^+(S)$ *but different sets of model functions* (c,η,N,X) *and* $(\tilde{c},\tilde{\eta},\tilde{N},\tilde{X})$ *satisfying Assumption 4.2.1. Then*

$$\begin{aligned}
\rho_F\left(\mu_t^{c,\eta,N,X},\mu_t^{\tilde{c},\tilde{\eta},\tilde{N},\tilde{X}}\right) &\leq C\int_0^t\left[\sup_{y\in S}\rho_F\left(\eta(\tau,y),\tilde{\eta}(\tau,y)\right)+\rho_F\left(N(\tau),\tilde{N}(\tau)\right)\right]\mathrm{d}\tau\\
&\quad+C\int_0^t\|c(\tau,\cdot)-\tilde{c}(\tau,\cdot)\|_\infty\,\mathrm{d}\tau+C\sup_{0\leq\tau_1\leq\tau_2\leq t}\|X(\tau_2,\tau_1,\cdot)-\tilde{X}(\tau_2,\tau_1,\cdot)\|_\infty,
\end{aligned}$$

i.e. the solutions are continuous with respect to model functions.

Combining the above results yields the following theorem.

Theorem 4.2.13 (Well-posedness of the linear model)**.** *Under Assumption 4.2.1, there exists a unique solution to the linear problem* (4.2.1) *with initial measure* $\mu_0\in\mathcal{M}^+(S)$*. Furthermore, the solution is continuous with respect to initial conditions and model functions.*

4.3 The generalised nonlinear model

Now that we solved the linear model (4.2.1), we go one step further and allow for nonlinearities, i.e. we consider the model functions of the form

$$\begin{aligned} &c : [0,T] \times S \times \mathcal{M}^+(S) \to \mathbb{R}, && \eta : [0,T] \times S \times \mathcal{M}^+(S) \to \mathcal{M}^+(S), \\ &N : [0,T] \times \mathcal{M}^+(S) \to \mathcal{M}^+(S), && X : [0,T] \times [0,T] \times S \times ([0,T] \to \mathcal{M}^+(S)) \to S. \end{aligned} \tag{4.3.1}$$

Remark 4.3.1. We require that the two-parameter family of Lipschitz maps X depends on a whole measure valued map $[0,T] \to \mathcal{M}^+(S)$ rather than on an individual measure in $\mathcal{M}^+(S)$. This assumption is based on the following observation: Recall that the map X is the generalisation of the flow of a vector field on $\mathbb{R}^d$ given by a nonlinear ODE of the form

$$\partial_t X_b(t;\tau,x) = b(t, X_b(t;\tau,x), \mu_t), \qquad X_b(\tau;\tau,x) = x, \tag{4.3.2}$$

where $b : [0,T] \times \mathbb{R}^d \times \mathcal{M}^+(\mathbb{R}^d) \to \mathbb{R}^d$ is a vector field and $\mu_\bullet : [0,T] \to \mathcal{M}^+(\mathbb{R}^d)$ is a measure-valued map. Integrating (4.3.2) from τ to t yields

$$X_b(t;\tau,x) = x + \int_\tau^t b(s, X_b(s;\tau,x), \mu_s)\,\mathrm{d}s,$$

so that $X_b(t,\tau,x)$ generally depends on values of μ_s for time arguments $\tau \le s \le t$.

We assume the following regularity for our model functions. Similar to the linear case, the generalised nonlinear model requires less regularity than the nonlinear Euclidean case; compare with Assumptions 3.5.1.

Assumptions 4.3.2 (Nonlinear model)**.** The model functions c, η and N satisfy the following regularity assumptions:

(i) For any $\mu \in \mathcal{M}^+(S)$, the model functions $c(\cdot,\cdot,\mu), \eta(\cdot,\cdot,\mu)$ and $N(\cdot,\mu)$ fulfill Assumption 4.2.1. Moreover, we have uniform bounds

$$\int_0^T \sup_{\mu \in \mathcal{M}^+(S)} \Big[\|c(\tau,\cdot,\mu)\|_\infty + \sup_{y \in S} \|\eta(\tau,y,\mu)\|_{BL^*} + \|N(\tau,\mu)\|_{BL^*} \Big]\,\mathrm{d}\tau < \infty, \tag{4.3.3}$$

and for any $R > 0$

$$\int_0^T \sup_{\|\mu\|_{BL^*} \le R} \Big[|c(\tau,\cdot,\mu)|_{\mathbf{Lip}} + \|\eta(\tau,\cdot,\mu)\|_{BL(S;\mathcal{M}^+)} \Big] \,\mathrm{d}\tau < \infty.$$

(ii) For all narrowly continuous maps $\mu_\bullet : [0,T] \to \mathcal{M}^+(S)$, $X(t,\tau,x,\mu_\bullet)$ is a two-parameter family of Lipschitz maps in t and τ. Moreover, for any R and any $\mu_\bullet$ with $\sup_{t\in[0,T]} \|\mu_t\|_{BL^*} \le R$, there exists a modulus of continuity $\omega_{X,R} : [0,\infty] \to [0,\infty]$ and a locally bounded function $L_{X,R} : \mathbb{R} \to \mathbb{R}$ satisfying

$$\begin{aligned} &\sup_{\tau\in\mathbb{R}} \sup_{x\in S} d(X(t_1,\tau,x,\mu_\bullet), X(t_2,\tau,x,\mu_\bullet)) \le \omega_{X,R}(|t_1-t_2|), \\ &d(X(t,\tau,x_1,\mu_\bullet), X(t,\tau,x_2,\mu_\bullet)) \le L_{X,R}(t-\tau)\, d(x_1,x_2), \qquad . \end{aligned} \tag{4.3.4}$$

In other words, X uniformly satisfies the properties of a two-parameter family of Lipschitz maps for each ball in the space of measures.

(iii) For any $R > 0$, there exists a function $L_{R,c} \in L^1(0,T)$ which is a.e. nonnegative and so that for $\|\mu\|_{BL^*}, \|\nu\|_{BL^*} \le R$

$$\|c(t,\cdot,\mu) - c(t,\cdot,\nu)\|_\infty \le L_{R,c}(t)\rho_F(\mu,\nu) \qquad \text{for a.e. } t \in [0,T].$$

(iv) For any $R > 0$, there exists a function $L_{R,\eta} \in L^1(0,T)$ which is a.e. nonnegative and so that for $\|\mu\|_{BL^*}, \|\nu\|_{BL^*} \le R$

$$\sup_{y\in S} \rho_F\left(\eta(t,y,\mu), \eta(t,y,\nu)\right) \le L_{R,\eta}(t)\, \rho_F(\mu,\nu) \qquad \text{for a.e. } t \in [0,T].$$

(v) For any $R > 0$, there exists a function $L_{R,N} \in L^1(0,T)$ which is a.e. nonnegative and so that for $\|\mu\|_{BL^*}, \|\nu\|_{BL^*} \le R$

$$\sup_{y\in S} \rho_F\left(N(t,\mu), N(t,\nu)\right) \le L_{R,N}(t)\, \rho_F(\mu,\nu) \qquad \text{for a.e. } t \in [0,T].$$

(vi) For any $R > 0$, there exists a function $L_{R,X} \in L^1(0,T)$ which is a.e. nonnegative and so that for $\sup_{t\in[0,T]} \|\mu_t\|_{BL^*}, \sup_{t\in[0,T]} \|\nu_t\|_{BL^*} \le R$

$$\|X(t_2,t_1,\cdot,\mu_\bullet) - X(t_2,t_1,\cdot,\nu_\bullet)\|_\infty \le \int_{t_1}^{t_2} L_{R,X}(\tau)\, \rho_F(\mu_\tau,\nu_\tau)\,\mathrm{d}\tau \qquad \forall t_1,t_2 \in [0,T].$$

With condition (4.3.3) we can bound the generalised solutions in the total variation norm so that all other bounds can be assumed to hold only locally.

We start with the notion of solution to model (4.3.1).

Definition 4.3.3. *We say that $\mu_\bullet : [0,T] \to \mathcal{M}^+(S)$ is a* ***(generalised) solution to the nonlinear structured population model on*** (S,d) *with initial measure $\mu_0 \in \mathcal{M}^+(S)$ and model functions c, η, N, X satisfying Assumption 4.3.2, if $\mu_\bullet$ is narrowly continuous and satisfies*

$$\begin{aligned} \mu_t &= X(t,0,\cdot,\mu_\bullet)_\# \left(\mu_0(\cdot) e^{\int_0^t c(s,\, X(s,0,\cdot,\mu_\bullet),\, \mu_s)\,\mathrm{d}s}\right) \\ &+ \int_0^t X(t,\tau,\cdot,\mu_\bullet)_\# \left(\int_S [\eta(\tau,y,\mu_\tau)(\cdot)]\,\mathrm{d}\mu_\tau(y) e^{\int_\tau^t c(s,\, X(s,\tau,\cdot,\mu_\bullet),\, \mu_s)\,\mathrm{d}s}\right)\mathrm{d}\tau \\ &+ \int_0^t X(t,\tau,\cdot,\mu_\bullet)_\# \left(N(\tau,\mu_\tau)(\cdot) e^{\int_\tau^t c(s,\, X(s,\tau,\cdot,\mu_\bullet),\, \mu_s)\,\mathrm{d}s}\right)\mathrm{d}\tau. \end{aligned} \tag{4.3.5}$$

Remark 4.3.4. For actually calculating integrals with measures of the form (4.3.5) we can again refer to Lemma 4.2.6 which naturally translates to the nonlinear case.

Similar to the Euclidean case, we show existence and uniqueness of the nonlinear model by a reduction to a linear version with fixed measure arguments. In contrast to Section 3.5, however, we do not conclude by a completeness argument of Cauchy sequences but rather invoke fixed point theory again.

Theorem 4.3.5. *Consider an initial measure $\mu_0 \in \mathcal{M}^+(S)$ and model functions (c,η,N,X) satisfying Assumptions 4.3.2. Then there exists a unique generalised solution of the nonlinear structured population model on (S,d) which is continuous with respect to initial conditions and model functions. More precisely, let (c,η,N,X) and $(\tilde{c},\tilde{\eta},\tilde{N},\tilde{X})$ be two sets of model functions which both satisfy Assumption 4.3.2 and let $\mu_t^{c,\eta,N,X}$ and $\mu_t^{\tilde{c},\tilde{\eta},\tilde{N},\tilde{X}}$ be the solutions to the corresponding nonlinear structured* *population models with initial measure μ_0. Then the following estimate holds*

$$\begin{aligned}\rho_F\big(\mu_t^{c,\eta,N,X},\, \mu_t^{\tilde c,\tilde\eta,\tilde N,\tilde X}\big) &\le C\int_0^t \sup_{\nu\in\mathcal{M}^+(S)} \|c(\tau,\cdot,\nu)-\tilde c(\tau,\cdot,\nu)\|_\infty \,\mathrm{d}\tau \\ &+ C\int_0^t \sup_{y\in S}\sup_{\nu\in\mathcal{M}^+(S)} \rho_F\left(\eta(\tau,y,\nu),\tilde\eta(\tau,y,\nu)\right)\mathrm{d}\tau \\ &+ C\int_0^t \sup_{\nu\in\mathcal{M}^+(S)} \rho_F\Big(N(\tau,\nu),\tilde N(\tau,\nu)\Big)\,\mathrm{d}\tau \\ &+ C \sup_{\nu_\bullet\in C^0([0,t];\,\mathcal{M}^+(S))}\ \sup_{0\le\tau_1\le\tau_2\le t} \|X(\tau_2,\tau_1,\cdot,\nu_\bullet)-\tilde X(\tau_2,\tau_1,\cdot,\nu_\bullet)\|_\infty.\end{aligned} \tag{4.3.6}$$

Similarly, if $\mu_t^{(1)}$ and $\mu_t^{(2)}$ are generalised solutions to the nonlinear structured population model with model functions (c,η,N,X) but different initial conditions $\mu_0^{(1)}$ and $\mu_0^{(2)}$, then there is a constant C_M such that

$$\rho_F(\mu_t^{(1)},\mu_t^{(2)}) \le C_M\, \rho_F(\mu_0^{(1)},\mu_0^{(2)}). \tag{4.3.7}$$

Proof. As already mentioned above, the key idea of the proof is to reduce the nonlinear model to a linear version with fixed the measure argument. Consider a measure valued map $\mu_\bullet$ in $C^0([0,T];\mathcal{M}^+(S))$ and define the operator $F\mu_\bullet := \nu_\bullet$, where $\nu_\bullet$ is the unique solution of the linear equation

$$\begin{aligned}\nu_t &= X(t,0,\cdot,\mu_\bullet)_\#\left(\mu_0(\cdot)e^{\int_0^t c(s,\,X(s,0,\cdot,\,\mu_\bullet),\mu_s)\,\mathrm{d}s}\right) \\ &+\int_0^t X(t,\tau,\cdot,\mu_\bullet)_\#\left(\int_S [\eta(\tau,y,\mu_\tau)(\cdot)]\,\mathrm{d}\nu_\tau(y)e^{\int_\tau^t c(s,\,X(s,\tau,\cdot,\,\mu_\bullet),\,\mu_s)\,\mathrm{d}s}\right)\mathrm{d}\tau \\ &+\int_0^t X(t,\tau,\cdot,\mu_\bullet)_\#\left(N(\tau,\mu_\tau)(\cdot)e^{\int_\tau^t c(s,\,X(s,\tau,\cdot,\,\mu_\bullet),\,\mu_s)\,\mathrm{d}s}\right)\mathrm{d}\tau.\end{aligned}$$

Existence and uniqueness of such a measure valued map $\nu_\bullet\in C^0([0,T];\mathcal{M}^1(S))$ follows from Theorem 4.2.13 if the adjusted model functions satisfy Assumptions 4.2.1. In particular, we need to check that $c(t,x,\mu_t)$, $\eta(t,x,\mu_t)(\cdot)$ and $N(t,\mu_t)(\cdot)$ are integrable with respect to time. By assumption, the map $\mu_\bullet$ is continuous and hence $\sup_{t\in[0,T]}\|\mu_t\|_{BL^*} =: R<\infty$. Thus, we conclude the integrability condition from the uniform bounds (4.3.3) in Assumptions 4.3.2. Similar to Lemma 4.2.8 in the linear case, we can show that $\nu_\bullet$ is in $C^0([0,T];\mathcal{M}^+(S))$ and uniformly bounded by Lemma 4.2.10. However, we need to guarantee that the iterative application of F preserves the uniform boundedness of the measure valued maps. To this end, we introduce the set

$$\mathcal{K}_C := \{\nu_\bullet \in C^0([0,T];\mathcal{M}^+(S)) \mid \nu_0 = \mu_0,\ \sup_{t\in[0,T]} \|\nu_t\|_{BL^*} \leq C\}.$$

Recall that the bounds from condition (4.3.3) hold uniformly and not only for measures μ with $\|\mu\|_{BL^*} \leq R$ for some $R > 0$. Applying these bounds to the uniform bound in Lemma 4.2.10 implies that we can choose C sufficiently large to achieve $F : \mathcal{K}_C \to \mathcal{K}_C$.

Next, we show that F is contractive. To this end, consider two generalised solutions $\mu_\bullet^{(1)}, \mu_\bullet^{(2)} \in C^0([0,T];\mathcal{M}^+(S))$ with $\mu_0^{(1)} = \mu_0^{(2)} = \mu_0$ and let $\nu_\bullet^{(1)} = F(\mu_\bullet^{(1)})_\bullet$ and $\nu_\bullet^{(2)} = F(\mu_\bullet^{(2)})_\bullet$. By Proposition 4.2.12 the generalised solutions are continuous with respect to model functions which leads to

$$\begin{aligned}
&\rho_F\left(\nu_t^{(1)}, \nu_t^{(2)}\right) \leq C_M \int_0^t \sup_{y\in S} \rho_F\left(\eta(\tau,y,\mu_\tau^{(1)}), \eta(\tau,y,\mu_\tau^{(2)})\right) \mathrm{d}\tau \\
&\quad + C_M \int_0^t \left[\rho_F\left(N(\tau,\mu_\tau^{(1)}), N(\tau,\mu_\tau^{(2)})\right) + \|c(\tau,\cdot,\mu_\tau^{(1)}) - c(\tau,\cdot,\mu_\tau^{(2)})\|_\infty\right] \mathrm{d}\tau \\
&\quad + C_M \sup_{0\leq\tau_1\leq\tau_2\leq t} \|X(\tau_2,\tau_1,\cdot,\mu_\bullet^{(1)}) - X(\tau_2,\tau_1,\cdot,\mu_\bullet^{(2)})\|_\infty.
\end{aligned}$$

We apply Assumptions 4.3.2 with $R = C$ and see

$$\begin{aligned}
&\rho_F\left(\nu_t^{(1)}, \nu_t^{(2)}\right) \leq C_M \int_0^t \left(L_{R,\eta}(\tau) + L_{R,N}(\tau) + L_{R,c}(\tau)\right) \rho_F(\mu_\tau^{(1)}, \mu_\tau^{(2)}) \,\mathrm{d}\tau \\
&\quad + C_M \sup_{0\leq\tau_1\leq\tau_2\leq t} \int_{\tau_1}^{\tau_2} L_{R,X}(\tau)\, \rho_F(\mu_\tau^{(1)}, \mu_\tau^{(2)}) \,\mathrm{d}\tau \\
&\leq C_M \sup_{0\leq\tau_1\leq\tau_2\leq t} \int_{\tau_1}^{\tau_2} L_R(\tau)\, \rho_F(\mu_\tau^{(1)}, \mu_\tau^{(2)}) \,\mathrm{d}\tau,
\end{aligned}$$

with the a.e nonnegative function $L_R = L_{R,\eta} + L_{R,N} + L_{R,c} + L_{R,X} \in L^1(0,T)$. We apply the generalised Bielecki norm (4.2.7) once more with some $\lambda > 0$ and $f = L_R$ and see that

$$\begin{aligned}
&\rho_{\lambda,L_R}(\nu_\bullet^{(1)}, \nu_\bullet^{(2)}) \leq \sup_{t\in[0,T]} \left[e^{-\lambda\int_0^t L_R(u)\,\mathrm{d}u} \rho_F(\nu_t^{(1)}, \nu_t^{(2)})\right] \\
&\leq C_M \sup_{t\in[0,T]} e^{-\lambda\int_0^t L_R(u)\,\mathrm{d}u} \sup_{0\leq\tau_1\leq\tau_2\leq t} \int_{\tau_1}^{\tau_2} L_R(\tau) \rho_F(\mu_\tau^{(1)}, \mu_\tau^{(2)}) \,\mathrm{d}\tau \\
&\leq C_M\, \rho_{\lambda,L_R}(\mu_\bullet^{(1)}, \mu_\bullet^{(2)}) \sup_{t\in[0,T]} e^{-\lambda\int_0^t L_R(u)\,\mathrm{d}u} \sup_{0\leq\tau_1\leq\tau_2\leq t} \int_{\tau_1}^{\tau_2} L_R(\tau)\, e^{\lambda\int_0^\tau L_R(u)\,\mathrm{d}u} \,\mathrm{d}\tau.
\end{aligned}$$

For $0 \leq \tau_1 \leq \tau_2 \leq t$ we can estimate the last factor as

$$\int_{\tau_1}^{\tau_2} L_R(\tau) e^{\lambda \int_0^\tau L_R(u)\,\mathrm{d}u}\,\mathrm{d}\tau = \frac{1}{\lambda}\left[e^{\lambda \int_0^{\tau_2} L_R(u)\,\mathrm{d}u} - e^{\lambda \int_0^{\tau_1} L_R(u)\,\mathrm{d}u}\right] \leq \frac{1}{\lambda}\left[e^{\lambda \int_0^{t} L_R(u)\,\mathrm{d}u} - 1\right],$$

so that we obtain the final estimate

$$\begin{aligned}\rho_{\lambda,L_R}(\nu_\bullet^{(1)},\nu_\bullet^{(2)}) &\leq \frac{C_M}{\lambda}\,\rho_{\lambda,L_R}(\mu_\bullet^{(1)},\mu_\bullet^{(2)}) \sup_{t\in[0,T]} e^{-\lambda\int_0^t L_R(u)\,\mathrm{d}u}\left[e^{\lambda\int_0^t L_R(u)\,\mathrm{d}u} - 1\right] \\ =&\frac{C_M}{\lambda}\,\rho_{\lambda,L_R}(\mu_\bullet^{(1)},\mu_\bullet^{(2)})\left[1 - e^{-\lambda\int_0^T L_R(u)\,\mathrm{d}u}\right] \leq \frac{C_M}{\lambda}\rho_{\lambda,L_R}(\mu_\bullet^{(1)},\mu_\bullet^{(2)}).\end{aligned}$$

In particular, F is contractive for $\lambda = 2C_M$ and thus we conclude existence and uniqueness by Banach Fixed Point Theorem.
We complete the proof with the continuity estimates (4.3.6) and (4.3.7). These follow directly from the corresponding results for the linear problem, i.e. Propositions 4.2.11 and 4.2.12. □

4.4 Model equivalence in the Euclidean case

We introduced the concept of generalised solutions to extend the structured population models to Polish metric spaces. In this section, we verify that in the $\mathbb{R}^d$ case the generalised solution and the measure solution to (3.5.1) are indeed equivalent. We choose the flow of the vector field b given by the unique solution of the ODE

$$\partial_t X_b(t;\tau,x,\mu_\bullet) = b(t, X_b(t;\tau,x,\mu_\bullet),\mu_t) \qquad X_b(\tau;\tau,x,\mu_\bullet) = x. \tag{4.4.1}$$

as two-parameter family of Lipschitz maps. Our first lemma shows that under mild assumptions this is actually a legit choice, i.e. that X_b satisfies (ii) and (vi) in Assumption 4.3.2 and is thus an explicit example for a two-parameter family of Lipschitz maps.

Assumptions 4.4.1 (Special case $\mathbb{R}^d$)**.** The assumptions on the model functions c, η and N remain unchanged, i.e. they still satisfy Assumptions 4.3.2. However, hypotheses (ii) and (vi) on X are replaced by the following assumptions on b : $[0,T]\times\mathbb{R}^d\times\mathcal{M}^+(\mathbb{R}^d)\to\mathbb{R}^d$:

(i) For any $\mu\in\mathcal{M}^+(\mathbb{R}^d)$, the function $b(\cdot,\cdot,\mu)$ is in $L^1((0,T);BL(\mathbb{R}^d;\mathbb{R}^d))$. Moreover, for any $R > 0$

$$\int_0^T \sup_{\|\mu\|_{BL^*} \leq R} \|b(\tau, \cdot, \mu)\|_{BL} \, \mathrm{d}\tau < \infty.$$

(ii) For any $R > 0$, there exists a function $L_{R,b} \in L^1(0,T)$ which is a.e. nonnegative and so that if $\mu_\bullet$, $\nu_\bullet$ are narrowly continuous and $\sup_{t\in[0,T]} \|\mu_t\|_{BL^*}, \|\nu_t\|_{BL^*} \leq R$, then

$$\|b(t, \cdot, \mu) - b(t, \cdot, \nu)\|_\infty \leq L_{R,b}(t)\rho_F(\mu, \nu) \qquad \text{for a.e. } t \in [0,T].$$

Note that in comparison to Assumptions 3.5.1 model function b needs slightly less regularity in the time argument. In particular, we only need to assume that $L_{R,b} \in L^1(0,T)$. This is consistent with the assumptions made on the other model functions in Assumptions 4.3.2.

Now we show that two-parameter families of Lipschitz maps indeed generalise the situation known for classical flows of vector fields.

Lemma 4.4.2. *Suppose that $b : [0,T] \times \mathbb{R}^d \times \mathcal{M}^+(\mathbb{R}^d) \to \mathbb{R}^d$ satisfies Assumption 4.4.1. Furthermore, let $\mu_\bullet$ be a narrowly continuous family in $\mathcal{M}^+(\mathbb{R}^d)$ and let $X_b(t;\tau,x,\mu_\bullet)$ be defined as the unique solution of the ODE* (4.4.1). *Then X_b satisfies (ii) and (vi) in Assumptions 4.3.2.*

Proof. We start with condition (ii) in Assumption 4.3.2, so let $R > 0$ and consider a narrowly continuous family $\mu_\bullet \subset \mathcal{M}^+(S)$ with $\sup_{t\in[0,T]} \|\mu_t\|_{BL^*} \leq R$. We compute

$$\begin{aligned}
&|X_b(t;\tau,x_1,\mu_\bullet) - X_b(t;\tau,x_2,\mu_\bullet)| \\
&\leq \int_\tau^t |b(s, X_b(s;\tau,x_1,\mu_\bullet),\mu_s) - b(s, X_b(s;\tau,x_2,\mu_\bullet),\mu_s)| \, \mathrm{d}s + |x_1 - x_2| \\
&\leq \int_\tau^t |b(s,\cdot,\mu_s)|_{\mathbf{Lip}} \, |X_b(s;\tau,x_1,\mu_\bullet) - X_b(s,\tau,x_2,\mu_\bullet)| \, \mathrm{d}s + |x_1 - x_2|
\end{aligned}$$

and thus we conclude

$$|X_b(t,\tau,x_1,\mu_\bullet) - X_b(t,\tau,x_2,\mu_\bullet)| \leq L_{X,R}(t-\tau) \, |x_1 - x_2|$$

with Gronwall's inequality and a suitable function $L_{X,R}$ given by

$$L_{X,R}(t) = \exp\left(\sup_{\substack{t_1,t_2\in[0,T] \\ |t_1-t_2|\le t}} \int_{t_1}^{t_2} \sup_{\|\mu\|_{BL^*}\le R} \|b(s,\cdot,\mu)\|_{BL}\, \mathrm{d}s \right).$$

To show the first estimate of (ii), let $t_2 > t_1$ and compute

$$\begin{aligned} |X_b(t_1;\tau,x,\mu_\bullet) - X_b(t_2;\tau,x,\mu_\bullet)| &\le \left| \int_{t_1}^{t_2} b(s, X_b(s;\tau,x,\mu_\bullet),\mu_s)\,\mathrm{d}s \right| \\ \le \int_{t_1}^{t_2} \sup_{x\in\mathbb{R}^d} |b(s,x,\mu_s)|\,\mathrm{d}s \le \sup_{t:\ t+|t_2-t_1|\le T} &\int_t^{t+|t_2-t_1|} \sup_{\|\mu\|_{BL^*}\le R} \|b(s,\cdot,\mu)\|_\infty \,\mathrm{d}s. \end{aligned}$$

By Assumptions 4.4.1, the function $s \mapsto \sup_{\|\mu\|_{BL^*}\le R} \|b(s,\cdot,\mu)\|_{BL}$ is in $L^1(0,T)$, so that b is uniformly integrable and thus

$$\omega_{X,R}(r) := \sup_{t:\ t+r\le T} \int_t^{t+r} \sup_{\|\mu\|_{BL^*}\le R} \|b(s,\cdot,\mu)\|_\infty \,\mathrm{d}s \to 0 \qquad (r\to 0).$$

In particular, $\omega_{X,R}(r)$ is a well-defined modulus of continuity and conditions (4.3.4) are fulfilled.

We are still left to check Assumption 4.3.2 (vi). So let $\mu_\bullet, \nu_\bullet \subset \mathcal{M}^+(\mathbb{R}^d)$ be two narrowly continuous families with $\sup_{t\in[0,T]} \|\mu_t\|_{BL^*}, \|\nu_t\|_{BL^*} \le R$. Then using (ii) and (i) of Assumptions 4.4.1 we compute

$$\begin{aligned} &|X_b(t;\tau,x,\mu_\bullet) - X_b(t;\tau,x,\nu_\bullet)| \\ &= |X_b(t;\tau,x,\mu_\bullet) - X_b(\tau;\tau,x,\mu_\bullet) + X_b(\tau;\tau,x,\nu_\bullet) - X_b(t;\tau,x,\nu_\bullet)| \\ &\le \int_\tau^t |b(s,X_b(s;\tau,x,\mu_\bullet),\mu_s) - b(s,X_b(s;\tau,x,\nu_\bullet),\nu_s)|\,\mathrm{d}s \\ &\le \int_\tau^t |b(s,X_b(s;\tau,x,\mu_\bullet),\mu_s) - b(s,X_b(s;\tau,x,\mu_\bullet),\nu_s)| \\ &\qquad + \int_\tau^t |b(s,X_b(s;\tau,x,\mu_\bullet),\nu_s) - b(s,X_b(s;\tau,x,\nu_\bullet),\nu_s)|\,\mathrm{d}s \\ &\le \int_\tau^t L_{R,b}(s)\rho_F(\mu_s,\nu_s)\,\mathrm{d}s \\ &\qquad + \int_\tau^t \left(\sup_{\|\nu\|_{BL^*}\le R} |b(s,\cdot,\nu)|_{\mathbf{Lip}} \right) |X_b(s;\tau,x,\mu_\bullet) - X_b(s;\tau,x,\nu_\bullet)|\,\mathrm{d}s. \end{aligned}$$

Applying Gronwall's inequality again concludes the proof. □

Theorem 4.4.3. *Suppose Assumptions 4.4.1 hold true. Let $\mu_0 \in \mathcal{M}^+(\mathbb{R}^d)$ and let $X_b(t,\tau,x,\mu_\bullet)$ be defined by* (4.4.1). *Then the generalised solution given by representation formula* (4.3.5) *satisfies the weak formulation* (3.5.2) *and is in particular a measure solution to* (3.5.1).

Proof. Using an approximation argument we can assume that the test functions are compactly supported in space. We refer to Remark 4.4.4 for the details. According to the density result Theorem D.5 in Ref. [44] we can further restrict the class of test functions to functions of the form $f(t,x) = \varphi(t)\psi(x)$ with $\varphi \in C^1([0,T]) \cap W^{1,\infty}([0,T])$ and $\psi \in C_c^1(\mathbb{R}^d)$.
We claim that the theorem will follow if we show that the map

$$[0,T] \ni t \mapsto \mathcal{F}(t) = \int_{\mathbb{R}^d} \psi(x)\,\mathrm{d}\mu_t(x)$$

is differentiable a.e. in t with derivative

$$\begin{aligned}\mathcal{F}'(t) = &\int_{\mathbb{R}^d} \nabla_x\psi(x)\cdot b(t,x,\mu_t) + \psi(x)\,c(t,x,\mu_t)\,\mathrm{d}\mu_t(x)\\ &+ \int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d}\psi(y)\,\mathrm{d}[\eta(t,x,\mu_t)]\,(y)\right)\mathrm{d}\mu_t(x) + \int_{\mathbb{R}^d}\psi(x)\,\mathrm{d}[N(t,\mu_t)]\,(x).\end{aligned} \tag{4.4.2}$$

Let us assume for the moment that (4.4.2) holds. In this case, Assumption 4.3.2 implies that $\mathcal{F}'$ is integrable and consequently $\mathcal{F} \in W^{1,1}([0,T])$, i.e. $\mathcal{F}$ is absolutely continuous. Due to the Fundamental Theorem of Calculus and the product rule we see

$$\begin{aligned}&\int_{\mathbb{R}^d} f(T,x)\,\mathrm{d}\mu_T(x) - \int_{\mathbb{R}^d} f(0,x)\,\mathrm{d}\mu_0(x) = \varphi(T)\mathcal{F}(T) - \varphi(0)\mathcal{F}(0)\\ &= \int_0^T \partial_t[\varphi(t)\mathcal{F}(t)]\,\mathrm{d}t = \int_0^T \partial_t\varphi(t)\mathcal{F}(t) + \varphi(t)\mathcal{F}'(t)\,\mathrm{d}t,\end{aligned}$$

which is exactly the right-hand side of (3.5.2) with test function $f(t,x) = \varphi(t)\psi(x)$.
We still have to show (4.4.2). To simplify the upcoming computations we introduce the following abbreviations

$$\mathcal{X}_{t,\tau}(x) := X_b(t;\tau,x,\mu_\bullet), \qquad \mathcal{E}_{t,\tau}(x) := e^{\int_\tau^t c(s,\mathcal{X}_{s,\tau}(x),\mu_s)\,\mathrm{d}s},$$

which brings (the nonlinear version of) (4.2.4) in the form

$$\begin{aligned}\int_{\mathbb{R}^d} \psi(x)\,\mathrm{d}\mu_t(x) &= \int_{\mathbb{R}^d} \psi(\mathcal{X}_{t,0}(x))\,\mathcal{E}_{t,0}(x)\,\mathrm{d}\mu_0(x)\\ &+ \int_0^t \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \psi(\mathcal{X}_{t,\tau}(y))\,\mathcal{E}_{t,\tau}(y)\,\mathrm{d}[\eta(\tau,x,\mu_\tau)](y)\,\mathrm{d}\mu_\tau(x)\,\mathrm{d}\tau \\ &+ \int_0^t \int_{\mathbb{R}^d} \psi(\mathcal{X}_{t,\tau}(x))\,\mathcal{E}_{t,\tau}(x)\,\mathrm{d}[N(\tau,\mu_\tau)](x)\,\mathrm{d}\tau.\end{aligned} \tag{4.4.3}$$

To prove formula (4.4.2) each term on the right- hand side of (4.4.3) has to be differentiated. We start with the first term and use (4.4.1) to compute

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathbb{R}^d} \psi(\mathcal{X}_{t,0}(x))\,\mathcal{E}_{t,0}(x)\,\mathrm{d}\mu_0(x)\right]\\ &= \int_{\mathbb{R}^d} [\nabla_x \psi(\mathcal{X}_{t,0}(x)) \cdot b(t,\mathcal{X}_{t,0}(x),\mu_t)]\,\mathcal{E}_{t,0}(x)\,\mathrm{d}\mu_0(x)\\ &\quad + \int_{\mathbb{R}^d} \psi(\mathcal{X}_{t,0}(x))\,\mathcal{E}_{t,0}(x)\,c(t,\mathcal{X}_{t,0}(x),\mu_t)\,\mathrm{d}\mu_0(x) =: A + B.\end{aligned}$$

The treatment of the second term of (4.4.3) requires differentiation under the Bochner integral so that we have to apply the Dominated Convergence Theorem for Bochner integrals (see e.g. Theorem H.7 in Ref. [44]) which yields

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_0^t \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \psi(\mathcal{X}_{t,\tau}(y))\,\mathcal{E}_{t,\tau}(y)\,\mathrm{d}[\eta(\tau,x,\mu_\tau)](y)\,\mathrm{d}\mu_\tau(x)\,\mathrm{d}\tau\right]\\ &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \psi(y)\,\mathrm{d}[\eta(t,x,\mu_t)](y)\,\mathrm{d}\mu_t(x)\\ &\quad + \int_0^t \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} [\nabla_y \psi(\mathcal{X}_{t,\tau}(y)) \cdot b(t,\mathcal{X}_{t,\tau}(y),\mu_t)]\;\mathcal{E}_{t,\tau}(y)\,\mathrm{d}[\eta(\tau,x,\mu_\tau)](y)\,\mathrm{d}\mu_\tau(x)\,\mathrm{d}\tau\\ &\quad + \int_0^t \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \psi(\mathcal{X}_{t,\tau}(y))\,\mathcal{E}_{t,\tau}(y)\,c(t,\mathcal{X}_{t,\tau}(y),\mu_t)\,\mathrm{d}[\eta(\tau,x,\mu_\tau)](y)\,\mathrm{d}\mu_\tau(x)\,\mathrm{d}\tau\\ &=: C + D + E.\end{aligned}$$

The third term in (4.4.3) can be computed analogously to be

$$
\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_0^t \int_{\mathbb{R}^d} \psi(\mathcal{X}_{t,\tau}(x))\, \mathcal{E}_{t,\tau}(x)\, \mathrm{d}[N(\tau,\mu_\tau)]\,(x)\, \mathrm{d}\tau\right] \\
&= \int_{\mathbb{R}^d} \psi(x)\, \mathrm{d}[N(t,\mu_t)](x) \\
&\quad + \int_0^t \int_{\mathbb{R}^d} \left[\nabla_x \psi(\mathcal{X}_{t,\tau}(x)) \cdot b(t, \mathcal{X}_{t,\tau}(x), \mu_t)\right]\, \mathcal{E}_{t,\tau}(x)\, \mathrm{d}[N(\tau,\mu_\tau)](x)\, \mathrm{d}\tau \\
&\quad + \int_0^t \int_{\mathbb{R}^d} \psi(\mathcal{X}_{t,\tau}(x))\, \mathcal{E}_{t,\tau}(x)\, c(t, \mathcal{X}_{t,\tau}(x), \mu_t)\, \mathrm{d}[N(\tau,\mu_\tau)](x)\, \mathrm{d}\tau \\
&=: F + G + H.
\end{aligned}
$$

We collect all terms involving $\nabla_x \psi \cdot b$ and use formula (4.2.4) which leads to the following simplification

$$
A + D + G = \int_{\mathbb{R}^d} \nabla_x \psi(x) \cdot b(t, x, \mu_t)\, \mathrm{d}\mu_t(x), \tag{4.4.4}
$$

and similarly for all terms with $\psi\, c$

$$
B + E + H = \int_{\mathbb{R}^d} \psi(x)\, c(t, x, \mu_t)\, \mathrm{d}\mu_t(x). \tag{4.4.5}
$$

Combining (4.4.4) and (4.4.5) with the remaining terms C and F leads to formula (4.4.2) as desired. $\square$

Remark 4.4.4. We claim that during the analysis of the PDE model (3.5.2) it is actually sufficient to consider test functions which are compactly supported in space. Let $\varphi \in BL([0,T] \times \mathbb{R}^d) \cap C^1([0,T] \times \mathbb{R}^d)$ be a test function and let $\xi \in C_c^\infty(\mathbb{R}^d)$ be a smooth cut-off function with $\xi \equiv 1$ on $B_1(0)$, $0 \leq \xi \leq 1$ and $\mathrm{supp}(\xi) \subseteq B_2(0)$. For $n \in \mathbb{N}$ we introduce the following sequence of functions

$$
\varphi_n(t,x) = \varphi(t,x)\, \xi\left(\frac{x}{n}\right),
$$

where each component of x is multiplied with the scalar $1/n$. By construction $\varphi_n \to \varphi$ as well as $\partial_t \varphi_n \to \partial_t \varphi$ for $n \to \infty$ pointwise and by Dominated Convergence also in L^1. Thus, the only term in the weak formulation (3.5.2) which needs special attention is the term involving $\nabla_x\varphi$. We compute for arbitrary $n \in \mathbb{N}$

$$
\begin{aligned}
&\left|\int_0^T \int_{\mathbb{R}^d} [\nabla_x\varphi_n(t,x) - \nabla_x\varphi(t,x)] \cdot b(t,x,\mu_t)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t\right| \\
\leq &\left|\int_0^T \int_{\mathbb{R}^d\setminus B_{2n}(0)} [\nabla_x\varphi_n(t,x) - \nabla_x\varphi(t,x)] \cdot b(t,x,\mu_t)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t\right| \\
&+ \left|\int_0^T \int_{B_{2n}(0)} [\nabla_x\varphi_n(t,x) - \nabla_x\varphi(t,x)] \cdot b(t,x,\mu_t)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t\right| \\
= &\left|\int_0^T \int_{\mathbb{R}^d\setminus B_{2n}(0)} \nabla_x\varphi(t,x) \cdot b(t,x,\mu_t)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t\right| \\
&+ \left|\int_0^T \int_{B_{2n}(0)} \left[\nabla_x\varphi(t,x)\,\xi\left(\frac{x}{n}\right) + \varphi(t,x)\nabla_x\xi\left(\frac{x}{n}\right)\frac{1}{n} - \nabla_x\varphi(t,x)\right] \cdot b(t,x,\mu_t)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t\right| \\
=:&A + B.
\end{aligned}
$$

As $\nabla_x\varphi$ is bounded and by Assumption 4.4.1 $b(\cdot,\cdot,\mu) \in L^1$ for all $\mu \in \mathcal{M}^+(\mathbb{R}^d)$ we can apply Dominated Convergence Theorem to see that $A \to 0$ for $n \to \infty$.
We continue with the term B

$$
\begin{aligned}
B \leq &\int_0^T \int_{B_{2n}(0)} \left|\left[\nabla_x\varphi(t,x)\,\xi\left(\frac{x}{n}\right) + \varphi(t,x)\nabla_x\xi\left(\frac{x}{n}\right)\frac{1}{n} - \nabla_x\varphi(t,x)\right] \cdot b(t,x,\mu_t)\right| \mathrm{d}\mu_t(x)\,\mathrm{d}t \\
\leq &\|\varphi\|_{W^{1,\infty}} \int_0^T \int_{B_{2n}(0)} \left|\nabla_x\xi\left(\frac{x}{n}\right)\frac{1}{n}\right| |b(t,x,\mu_t)|\,\mathrm{d}\mu_t(x)\,\mathrm{d}t \\
&+ \|\varphi\|_{W^{1,\infty}} \int_0^T \int_{B_{2n}(0)} \left|\xi\left(\frac{x}{n}\right) - 1\right| |b(t,x,\mu_t)|\,\mathrm{d}\mu_t(x)\,\mathrm{d}t \\
=:&B_1 + B_2.
\end{aligned}
$$

For the first term B_1 we note that $\nabla_x\xi$ is bounded for all $x \in \mathbb{R}^d$. Furthermore, narrow continuity of the generalised solution $\mu_\bullet$ implies $\sup_{t\in[0,T]}\|\mu_t\|_{BL^*} =: R < \infty$. Thus, we compute

$$
\begin{aligned}
B_1 \leq &\frac{1}{n}\|\nabla_x\xi\|_\infty\|\varphi\|_{W^{1,\infty}} \int_0^T \|b(t,\cdot,\mu_t)\|_{BL}\|\mu_t\|_{BL^*}\,\mathrm{d}t \\
\leq &\frac{C}{n}\int_0^T \sup_{\|\mu\|_{BL^*}\leq R} \|b(t,\cdot,\mu)\|_{BL}\,\mathrm{d}t \to 0 \qquad (n\to\infty)
\end{aligned}
$$

by Assumption 4.4.1. Concerning B_2 we see that $\xi(\cdot/n) \equiv 1$ on $B_n(0)$, so that

$$
\begin{aligned}
B_2 \leq & \|\varphi\|_{W^{1,\infty}} \int_0^T \int_{B_{2n}\setminus B_n(0)} \left|\xi\left(\frac{x}{n}\right) - 1\right| |b(t,x,\mu_t)| \,\mathrm{d}\mu_t(x)\,\mathrm{d}t \\
\leq & \|\varphi\|_{W^{1,\infty}} \int_0^T \sup_{\|\mu\|_{BL^*}\leq R} \|b(t,\cdot,\mu)\|_{BL}\, \mu_t\left(B_{2n}\setminus B_n(0)\right)\mathrm{d}t \\
\leq & \|\varphi\|_{W^{1,\infty}} \sup_{t\in[0,T]} \mu_t\left(B_{2n}\setminus B_n(0)\right) \int_0^T \sup_{\|\mu\|_{BL^*}\leq R} \|b(t,\cdot,\mu)\|_{BL}\,\mathrm{d}t \\
= & C \sup_{t\in[0,T]} \mu_t\left(B_{2n}\setminus B_n(0)\right).
\end{aligned}
$$

In particular, $B_2 \to 0$ if $\sup_{t\in[0,T]} \mu_t\left(B_{2n}\setminus B_n(0)\right) \to 0$ for $n\to\infty$.
To see this, we first remember that the the map $t\mapsto \mu_t$ is narrowly continuous and thus the set $\mathcal{F} = \{\mu_t \mid t\in[0,T]\}$ is closed with respect to narrow convergence. We furthermore claim that $\mathcal{F}$ is even sequentially compact with respect to narrow convergence. Indeed, let $(\mu_{t_n})_{n\in\mathbb{N}} \subset \mathcal{F}$ be a sequence. By compactness the sequence of time points $(t_n)_{n\in\mathbb{N}} \subset [0,T]$ has a converging subsequence $\left(\tilde{t}_n\right)_{n\in\mathbb{N}}$ with $\tilde{t}_n \to \tilde{t} \in [0,T]$. Thus, by narrow continuity $\mu_{\tilde{t}_n} \to \mu_{\tilde{t}} \in \mathcal{F}$ and $\mathcal{F}$ is sequentially compact.
Now the Theorem of Prokhorov 2.1.3 implies that $\mathcal{F}$ is tight so that clearly

$$
\sup_{t\in[0,T]} \mu_t\left(B_{2n}\setminus B_n(0)\right) \to 0 \qquad (n\to\infty)
$$

as desired. This yields $B_2 \to 0$ and in particular

$$
\begin{aligned}
&\left|\int_0^T \int_{\mathbb{R}^d} \left[\nabla_x\varphi_n(t,x) - \nabla_x\varphi(t,x)\right]\cdot b(t,x,\mu_t)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t\right| \\
\leq & A + B_1 + B_2 \to 0 \quad (n\to\infty).
\end{aligned}
$$

Hence, the approximation is indeed valid in the limit.

We conclude this subsection with the missing uniqueness to the nonlinear PDE model (3.5.1). Together with Theorem 4.4.3 this will yield the desired equivalence of both the PDE and the generalised model. Before we start, we have to establish a regularity result.

Lemma 4.4.5. *Under Assumption 4.4.1 any measure solution $\mu_\bullet$ to model* (3.5.1) *induces an absolutely continuous map $[0,T] \ni t \mapsto \mu_t$, i.e. there exists a nonnegative* *function $L_\mu \in L^1(0,T)$ such that for any $0 \le \tau \le t \le T$*

$$\rho_F(\mu_\tau, \mu_t) \le \int_\tau^t L_\mu(s)\,\mathrm{d}s. \tag{4.4.6}$$

Furthermore, if $\mu_\bullet$ and $\tilde{\mu}_\bullet$ are two measure solution to (3.5.1), *then the map $[0,T] \ni t \mapsto \rho_F(\mu_t, \tilde{\mu}_t)$ is also absolutely continuous.*

Proof. The proof is similar to [44, Lemma 2.25] but adapted to the case $\mathbb{R}^d$.
For $\psi \in BL(\mathbb{R}^d) \cap C^1(\mathbb{R}^d)$, $0 < s_1 < s_2 < T$ and $\varepsilon > 0$ we use (3.5.2) with the test function

$$\varphi_\varepsilon(t,x) = \begin{cases} 0 & \text{if } t \in [0, s_1 - \varepsilon] \\ \left(\frac{t-s_1}{\varepsilon} + 1\right)\psi(x) & \text{if } t \in [s_1 - \varepsilon, s_1] \\ \psi(x) & \text{if } t \in [s_1, s_2] \\ \left(\frac{s_2-t}{\varepsilon} + 1\right)\psi(x) & \text{if } t \in [s_2, s_2 + \varepsilon] \\ 0 & \text{if } t \in [s_2 + \varepsilon, T] \end{cases}, \tag{4.4.7}$$

and obtain

$$\begin{aligned} &\int_{\mathbb{R}^d} \varphi_\varepsilon(T,x)\,\mathrm{d}\mu_T(x) - \int_{\mathbb{R}^d} \varphi_\varepsilon(0,x)\,\mathrm{d}\mu_0(x) = \int_0^T \int_{\mathbb{R}^d} \partial_t \varphi_\varepsilon(t,x)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t \\ &\quad + \int_0^T \int_{\mathbb{R}^d} \big(\nabla_x \varphi_\varepsilon(t,x) \cdot b(t,x,\mu_t) + \varphi_\varepsilon(t,x)\, c(t,x,\mu_t)\big)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t \\ &\quad + \int_0^T \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \varphi_\varepsilon(t,y)\,\mathrm{d}[\eta(t,x,\mu_t)]\,(y) \right) \mathrm{d}\mu_t(x)\,\mathrm{d}t + \int_0^T \int_{\mathbb{R}^d} \varphi_\varepsilon(t,x)\,\mathrm{d}[N(t,\mu_t)](x)\,\mathrm{d}t, \end{aligned} \tag{4.4.8}$$

We first note that by construction for ε small enough, we have

$$\int_{\mathbb{R}^d} \varphi_\varepsilon(T,x)\,\mathrm{d}\mu_T(x) = \int_{\mathbb{R}^d} \varphi_\varepsilon(0,x)\,\mathrm{d}\mu_0(x) = 0.$$

Furthermore,

$$
\begin{aligned}
&\int_0^T \int_{\mathbb{R}^d} \partial_t \varphi_\varepsilon(t,x)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t\\
=&\frac{1}{\varepsilon}\int_{s_1-\varepsilon}^{s_1}\int_{\mathbb{R}^d}\psi(x)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t - \frac{1}{\varepsilon}\int_{s_2}^{s_2+\varepsilon}\int_{\mathbb{R}^d}\psi(x)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t\\
\to&\int_{\mathbb{R}^d}\psi(x)\,\mathrm{d}\mu_{s_1}(x) - \int_{\mathbb{R}^d}\psi(x)\,\mathrm{d}\mu_{s_2}(x) \qquad (\varepsilon\to 0)
\end{aligned}
$$

by Lebesgue's Differentiation Theorem and continuity of the map $s \mapsto \int_{\mathbb{R}^d}\psi(x)\,\mathrm{d}\mu_s$. The other terms in (4.4.8) can be treated with Dominated Convergence Theorem as $\sup_{t\in[0,T]}\|\mu_t\|_{BL^*} =: R < \infty$ by narrow continuity. As an example, we treat the second term

$$
\begin{aligned}
&\int_0^T \int_{\mathbb{R}^d} \nabla_x \varphi_\varepsilon(t,x)\cdot b(t,x,\mu_t)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t\\
=&\int_{s_1-\varepsilon}^{s_1}\int_{\mathbb{R}^d}\frac{t-s_1}{\varepsilon}\nabla_x\varphi(t,x)\cdot b(t,x,\mu_t)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t + \int_{s_1}^{s_2}\int_{\mathbb{R}^d}\nabla_x\varphi(t,x)\cdot b(t,x,\mu_t)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t\\
&+\int_{s_2}^{s_2+\varepsilon}\int_{\mathbb{R}^d}\frac{s_2-t}{\varepsilon}\nabla_x\varphi(t,x)\cdot b(t,x,\mu_t)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t\\
=:&A+B+C.
\end{aligned}
$$

We apply Lebesgue's Differentiation Theorem

$$
\begin{aligned}
A =&\frac{1}{\varepsilon}\int_{s_1-\varepsilon}^{s_1}(s_1-t)\int_{\mathbb{R}^d}\nabla_x\varphi(t,x)\cdot b(t,x,\mu_t)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t\\
\to&(s_1-s_1)\int_{\mathbb{R}^d}\nabla_x\varphi(t,x)\cdot b(t,x,\mu_t)\,\mathrm{d}\mu_{s_1}(x) = 0 \qquad (\varepsilon\to 0).
\end{aligned}
$$

A similar computation shows $C \to 0$ so that we conclude for $\varepsilon \to 0$

$$
\int_0^T \int_{\mathbb{R}^d} \nabla_x \varphi_\varepsilon(t,x)\cdot b(t,x,\mu_t)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t \to \int_{s_1}^{s_2}\int_{\mathbb{R}^d}\nabla_x\varphi(t,x)\cdot b(t,x,\mu_t)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t.
$$

Combining the above computations, we see that sending $\varepsilon \to 0$ in (4.4.8) leads to

$$\begin{aligned}
&\int_{\mathbb{R}^d} \psi(x)\,\mathrm{d}\mu_{s_2}(x) - \int_{\mathbb{R}^d} \psi(x)\,\mathrm{d}\mu_{s_1}(x)\\
&= \int_{s_1}^{s_2} \int_{\mathbb{R}^d} \big(\nabla_x \psi(x)\cdot b(t,x,\mu_t) + \psi(x)\, c(t,x,\mu_t)\big)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t\\
&\quad + \int_{s_1}^{s_2} \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \psi(y)\,\mathrm{d}[\eta(t,x,\mu_t)]\,(y)\right) \mathrm{d}\mu_t(x)\,\mathrm{d}t + \int_{s_1}^{s_2} \int_{\mathbb{R}^d} \psi(x)\,\mathrm{d}[N(t,\mu_t)](x)\,\mathrm{d}t.
\end{aligned} \tag{4.4.9}$$

By Assumption 4.4.1 we have a bound

$$\begin{aligned}
&\int_{s_1}^{s_2} \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \psi(y)\,\mathrm{d}[\eta(t,x,\mu_t)]\,(y)\right) \mathrm{d}\mu_t(x)\,\mathrm{d}t\\
\leq & \int_{s_1}^{s_2} \|\psi\|_\infty \sup_{x\in\mathbb{R}^d} \|\eta(t,x,\mu_t)\|_{BL^*} \|\mu_t\|_{BL^*}\,\mathrm{d}t\\
\leq & R\,\|\psi\|_{BL} \int_{s_1}^{s_2} \sup_{\|\mu\|_{BL^*}\leq R}\, \sup_{x\in\mathbb{R}^d} \|\eta(t,x,\mu)\|_{BL^*}\,\mathrm{d}t.
\end{aligned}$$

Now, similar computations can be applied to the other terms appearing on the (RHS) of (4.4.9). Therefore, taking supremum over all $\psi \in BL(\mathbb{R}^d)$ with $\|\psi\|_{BL} \leq 1$ we prove (4.4.6). An analogous argument works when $s_1 = 0$ or $s_2 = T$.

Now consider two measure solutions $\mu_\bullet, \tilde{\mu}_\bullet$ to (3.5.1) and let $t > \tau$. Then we compute by triangle inequality and with (4.4.6)

$$\begin{aligned}
\rho_F(\mu_t, \tilde{\mu}_t) \leq & \rho_F(\mu_\tau, \tilde{\mu}_\tau) + \int_\tau^t L_\mu(s)\,\mathrm{d}s + \int_\tau^t L_{\tilde{\mu}}(s)\,\mathrm{d}s,\\
\rho_F(\mu_\tau, \tilde{\mu}_\tau) \leq & \rho_F(\mu_t, \tilde{\mu}_t) + \int_\tau^t L_\mu(s)\,\mathrm{d}s + \int_\tau^t L_{\tilde{\mu}}(s)\,\mathrm{d}s,
\end{aligned}$$

which can be combined to

$$|\rho_F(\mu_t, \tilde{\mu}_t) - \rho_F(\mu_\tau, \tilde{\mu}_\tau)| \leq \int_\tau^t \left(L_\mu(s) + L_{\tilde{\mu}}(s)\right)\mathrm{d}s.$$

In particular, the map $t \mapsto \rho_F(\mu_t, \tilde{\mu}_t)$ is absolutely continuous. □

We are now in the position to prove uniqueness of solutions to model (3.5.1).

Theorem 4.4.6. *Under Assumptions 3.5.1 (which are slightly stronger than Assumptions 4.4.1) measure solution to model* (3.5.1) *with initial condition* $\mu_0 \in$

$\mathcal{M}^+(\mathbb{R}^d)$ are unique.

Proof. Let $\mu_\bullet^{(1)}$ and $\mu_\bullet^{(2)}$ be two measure solutions to (3.5.1) with initial condition μ_0. By carefully modifying the model functions to

$$\begin{aligned} b^{(1)}(t,x) &= b\left(t,x,\mu_t^{(1)}\right), & c^{(1)}(t,x) &= c\left(t,x,\mu_t^{(1)}\right), \\ \eta^{(1)}(t,x) &= \eta\left(t,x,\mu_t^{(1)}\right), & N^{(1)}(t) &= N\left(t,\mu_t^{(1)}\right) \end{aligned}$$

we see that $\mu_\bullet^{(1)}$ can be interpreted as a measure solution to a linear model and similarly for $\mu_\bullet^{(2)}$. Due to the already established uniqueness in the linear case, see Lemma 3.4.5, together with Theorem 4.4.3, $\mu_\bullet^{(1)}$ and $\mu_\bullet^{(2)}$ are also (generalised) solutions to representation formula (4.2.1) with model functions $c^{(i)}, \eta^{(i)}, N^{(i)}$ and Lipschitz maps $X^{(i)}$ given by the flow of vector fields $b^{(i)}$, $i = 1, 2$.
Furthermore, according to Proposition 4.2.12 the solutions are continuous with respect to model functions so that

$$\begin{aligned} \rho_F\left(\mu_t^{(1)}, \mu_t^{(2)}\right) \leq & C \int_0^t \sup_{x\in\mathbb{R}^d} \rho_F\left(\eta\left(\tau,x,\mu_t^{(1)}\right), \eta\left(\tau,x,\mu_t^{(2)}\right)\right) \mathrm{d}\tau \\ & + C \int_0^t \rho_F\left(N\left(\tau,\mu_t^{(1)}\right), N\left(\tau,\mu_t^{(2)}\right)\right) \mathrm{d}\tau \\ & + C \int_0^t \left\| c\left(\tau,\cdot,\mu_t^{(1)}\right) - c\left(\tau,\cdot,\mu_t^{(2)}\right)\right\|_\infty \mathrm{d}\tau \\ & + C \sup_{0\leq\tau_1\leq\tau_2\leq t} \left\| X^{(1)}(\tau_2;\tau_1,\cdot) - X^{(2)}(\tau_2;\tau_1,\cdot)\right\|_\infty . \end{aligned}$$

We take a closer look at the last term and notice it can be simplified to

$$\begin{aligned} & C \sup_{0\leq\tau_1\leq\tau_2\leq t} \left\| X^{(1)}(\tau_2;\tau_1,\cdot) - X^{(2)}(\tau_2;\tau_1,\cdot)\right\|_\infty \\ \leq & C \sup_{0\leq\tau_1\leq\tau_2\leq t} \left\| \int_{\tau_1}^{\tau_2} b^{(1)}\left(\tau, X^{(1)}(\tau;\tau_1,\cdot)\right) - b^{(2)}\left(\tau, X^{(2)}(\tau;\tau_1,\cdot)\right) \mathrm{d}\tau \right\|_\infty \\ \leq & C \int_0^t \left\| b\left(\tau,\cdot,\mu_t^{(1)}\right) - b\left(\tau,\cdot,\mu_t^{(2)}\right)\right\|_\infty \mathrm{d}\tau. \end{aligned}$$

Applying the bounds from Assumption 4.4.1 hence implies

$$\rho_F\left(\mu_t^{(1)}, \mu_t^{(2)}\right) \leq C \int_0^t L(\tau)\, \rho_F\left(\mu_\tau^{(1)}, \mu_\tau^{(2)}\right) \mathrm{d}\tau$$

for some $L \in L^1(0,T)$, so that

$$\rho_F\left(\mu_t^{(1)}, \mu_t^{(2)}\right) = 0 \tag{4.4.10}$$

for a.e. $t \in [0,T]$ by Gronwall's inequality. As $t \mapsto \rho_F(\mu_t^{(1)}, \mu_t^{(2)})$ is continuous by Lemma 4.4.5 we conclude that (4.4.10) holds for all $t \in [0,T]$, establishing uniqueness and thus finishing the proof. $\square$

We conclude this chapter with a discussion on the validity of the implicit notion of solution in Definition 4.2.3.

Remark 4.4.7. If we insert the implicit integral representation (3.3.3) of $\varphi_{\psi,t}$ into the defining equation (3.4.2) for μ_t , we get for any $\psi \in BL(\mathbb{R}^d)$

$$\begin{aligned}
&\int_{\mathbb{R}^d} \psi(x)\,\mathrm{d}\mu_t(x) = \int_{\mathbb{R}^d} \varphi_{\psi,t}(0,x)\,\mathrm{d}\mu_0(x) + \int_0^t \int_{\mathbb{R}^d} \varphi_{\psi,t}(\tau,x)\,\mathrm{d}[N(\tau)](x)\,\mathrm{d}\tau \\
=&\int_{\mathbb{R}^d} \psi(X_b(t;0,x)) e^{\int_0^t c(r,X_b(r;0,x))\,\mathrm{d}r}\,\mathrm{d}\mu_0(x) \\
&+ \int_{\mathbb{R}^d}\int_0^t\int_{\mathbb{R}^d} \varphi_{\psi,t}(s,y)\,\mathrm{d}[\eta(s,X_b(s;0,x))](y) e^{\int_0^s c(r,X_b(r;0,x))\,\mathrm{d}r}\,\mathrm{d}s\,\mathrm{d}\mu_0(x) \\
&+ \int_0^t\int_{\mathbb{R}^d} \psi(X_b(t;\tau,x)) e^{\int_\tau^t c(r,X_b(r;\tau,x))\,\mathrm{d}r}\,\mathrm{d}[N(\tau)](x)\,\mathrm{d}\tau \\
&+ \int_0^t\int_{\mathbb{R}^d}\int_\tau^t\int_{\mathbb{R}^d} \varphi_{\psi,t}(s,y)\,\mathrm{d}[\eta(s,X_b(s;\tau,x))](y) e^{\int_\tau^s c(r,X_b(r;\tau,x))\,\mathrm{d}r}\,\mathrm{d}s\,\mathrm{d}[N(\tau)](x)\,\mathrm{d}\tau \\
=&A + B + C + D.
\end{aligned}$$

So if $\eta = 0$, this representation coincides with the implicit solution formula (4.2.4), i.e. the terms A and C appear in both cases. To ensure consistency of both representations in the presence of η, we would need that

$$\begin{aligned}
&B + D \\
=&\int_{\mathbb{R}^d}\int_0^t\int_{\mathbb{R}^d} \varphi_{\psi,t}(s,y)\,\mathrm{d}[\eta(s,X_b(s;0,x))](y) e^{\int_0^s c(r,X_b(r;0,x))\,\mathrm{d}r}\,\mathrm{d}s\,\mathrm{d}\mu_0(x) \\
&+ \int_0^t\int_{\mathbb{R}^d}\int_\tau^t\int_{\mathbb{R}^d} \varphi_{\psi,t}(s,y)\,\mathrm{d}[\eta(s,X_b(s;\tau,x))](y) e^{\int_\tau^s c(r,X_b(r;\tau,x))\,\mathrm{d}r}\,\mathrm{d}s\,\mathrm{d}[N(\tau)](x)\,\mathrm{d}\tau \\
\overset{!}{=}&\int_0^t\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \psi(X_b(t;\tau,y)) e^{\int_\tau^t c(s,X_b(s;\tau,y))\,\mathrm{d}s}\,\mathrm{d}[\eta(\tau,x)](y)\,\mathrm{d}\mu_\tau(x)\,\mathrm{d}\tau.
\end{aligned} \tag{4.4.11}$$

However, proving this identity turns out to be difficult, which admits two distinct

interpretations. Should (4.4.11) be true, then formula (4.2.4) or rather (4.2.1) indeed reflects all occurring effects which influence the behaviour of the solution μ_t. In this case, the proof of (4.4.11) presumably requires a change of variable formula in the spirit of (3.4.2) for general functions f instead of $\varphi_{\psi,t}$ on the right-hand side. Unfortunately, such a result is beyond the scope of this thesis.
Alternatively, it could also be that identity (4.4.11) fails to hold since formula (4.2.1) does not capture all relevant interactions of the model functions. In particular, term D contains an integral over η as well as over N. This could indicate that, apart from the individual effects of the mutation η (term B) and the state-independent influx N (term C), there is also an additional non-separable interaction effect between both terms. This is not entirely unlikely, as a similar behaviour can be observed for the flow X_b and the growth term c, which appear in all summands $A - D$. So in this case, formula (4.2.1) would have to be adjusted accordingly. However, the fact that the generalised solution coincides with the measure solution in the Euclidean case suggests otherwise.

5 Future perspectives and connection to other publications

In this chapter, we demonstrate how the developed theory can affect current research and point our future perspectives for further studies. As a start, we outline several model applications that can fully exploit the setting of non-Euclidean state spaces. More precisely, we present two examples that require a Polish metric space since they are inherently infinite dimensional as well as an additional example using a proper metric space. The latter is interesting in its own right due to its importance in differential geometry. Since the focus of this thesis is on theory and not on modelling, we only sketch the ideas without discussing explicit model functions. In the second part, we show how we can incorporate the concept of measure differential equations developed by Piccoli and the class of coagulation-fragmentation models into our framework. For the latter, we will need a minor adaptation of our theory, but are rewarded with a significant extension of our theory beyond structured population models. We conclude this chapter with a comparison of our flat metric to a transport type distance introduced by Fournier and Perthame.

5.1 Model examples motivated by applications

Example 5.1.1. One promising application of the new setting are structured population models that are based on functions or its trajectories instead of single individuals or population densities. In such a scenario, one would consider metric spaces of the form

$$(S, d) = \left(C^0([0, T]; \mathcal{X}), \|\cdot\|_\infty\right), \tag{5.1.1}$$

with some target space $\mathcal{X}$, e.g. $\mathbb{R}^d$ or $\mathbb{N}_0^d$. According to basic functional analysis, the space (S, d) is both separable and complete but not proper. If we want to study not only continuous but also discontinuous trajectories, we can simply consider the space $S = L^p([0, T]; \mathcal{X})$ instead of (5.1.1).

Such a modelling approach can be used, among other things, for inferring developmental trajectories of cellular differentiation from single-cell RNA-sequencing data (scRNA-seq data). The current biological consensus is that during differentiation an individual cell follows a continuous path in a *gene expression space* $\mathcal{X}$, the so-called *individual cell trajectory*. Together, all these trajectories form a metric space of the form (5.1.1). However, due to heterogeneity, cells of a particular cell type do not necessarily have a common developmental trajectory, but rather follow a *cell type specific trajectory distribution* $\mu \in \mathcal{M}^+(C^0([0,T];\mathcal{X}))$. This trajectory distribution can be expected to change as a result of biological processes, external influences or rapid epigenetic mutations [27].

Example 5.1.2. In this example we are interested in the pseudotemporal development of a specific gene, see also the related approach [124]. Thus, we modify the setting of the previous example by taking the average trajectory of a given cell type instead of the cell type specific trajectory distribution. The resulting *pseudotime trajectory* $\gamma : [0,T] \to \mathcal{X}$ hence roughly describes the differentiation process of the cells with the given cell type. Along the whole developmental path γ, the intensity of the expressed gene varies, e.g. represented by a change in absolute gene counts. Due to inherent cell heterogeneities one would however rather expect a gene expression distribution among all cells, $\mu_t \in \mathcal{M}^+(S)$ with $S = \mathbb{N}_0$ or $S = \mathbb{R}^+$.
If we want to consider more than one gene following the trajectory γ, there is a distribution $\mu_t^i \in \mathcal{M}^+(S)$ for each pseudotime point $t \in [0,T]$ and each gene $i = 1, ..., N$. In this situation, the evolution of all gene distributions along the trajectory γ can be represented by a measure on $\mathcal{M}^+(S)$, e.g. in the discrete case here by a measure of the form

$$M_t^N = \sum_{i=1}^{N} \alpha_i(t) D_{\mu_t^i} \in \mathcal{M}^+(\mathcal{M}^+(S)),$$

where $D_\mu \in \mathcal{M}^+(\mathcal{M}^+(S))$ denotes the Dirac measure on $\mathcal{M}^+(S)$ located in μ. More elaborate measures on $\mathcal{M}^+(S)$, such as measures without atoms, could occur in the limit $N \to \infty$. Although it is possible to model such objects by products of measure spaces, the number of genes N may be very large and thus it might be convenient to consider the limit of M_t^N in $\mathcal{M}^+(\mathcal{M}^+(S))$ as $N \to \infty$.

Our last example stems from the field of differential geometry. Although it cannot make full use of the new framework in its complete generality, it is at least set on a proper metric space. We refer to Chapter 6 in Ref. [101] for precise definitions and further information on Riemannian geometry.

Example 5.1.3. For a finite dimensional smooth manifold S consider a Riemannian metric g such that the pair (S, g) is a connected and geodesically complete Riemannian manifold. Well-known examples of the latter include the torus, the sphere or the hyperbolic space. A natural metric on S is given by the corresponding Riemannian distance function d_R, which measures the length of the shortest path between two points. According to the Theorem of Hopf-Rinow [89], geodesical completeness implies that the metric space (S, d_R) is proper (see Remark 2.3.5). This enables the formulation of structured population models on (S, d_R), such as those appearing in [21, 122, 130].

Remark 5.1.4. In addition to the above examples, our framework is also suitable for operating on discrete structures which are inherently Polish spaces if equipped with a suitable metric. For example, Refs. [107, 111] consider structured population models on graphs to investigate mutual cooperation. Similarly, in the field of image processing and high-dimensional data analysis, graph or network-like structures are very common as graphs provide a suitable ambient space for representing or classifying data [12, 73, 154]. In this setting, PDEs can be applied to simplify data processing or support semi-supervised learning approaches [19, 49, 109]. At this point we especially highlight the papers [50, 51] that study nonlocal measure-valued continuity equations on graphs, although under the total variation norm.

5.2 Connection to measure differential equations

In this section we turn to an alternative but related modelling approach based on so called measure differential equations (MDEs). MDEs have been introduced by Piccoli for pedestrian flows in terms of a conservative transport equation [119],

$$\dot{\mu}_t = V[\mu_t], \tag{5.2.1}$$

and thus provide a generalisation of the concept of ordinary differential equations to spaces of measures. In this approach, the evolution of a measure μ is governed by a **measure vector field** (MVF) $V[\mu]$ which is a measure on $\mathbb{R}^d \times \mathbb{R}^d$. Its first coordinate represents spatial position x and the second denotes admissible values of velocity v. The measure $V[\mu]$ has marginal μ on the first coordinate, i.e. it satisfies $\pi^1_\# V[\mu] = \mu$ for all $\mu \in \mathcal{M}^+(\mathbb{R}^d)$. Here π^1 denotes the projection to the first (spatial) coordinate and the subscript # denotes the push-forward operator (see Definition 2.1.5). Roughly speaking, if (x, v) belongs to the support of $V[\mu]$, μ at position x evolves with velocity v. We refer to [119, Section 7.1] for examples of measure vector fields. In the follow-up paper [123] by Piccoli and Rossi equation (5.2.1) has been extended by a nonnegative source term $s : \mathcal{M}^+(\mathbb{R}^d) \to \mathcal{M}^+(\mathbb{R}^d)$

$$\dot{\mu}_t = V[\mu_t] \oplus s[\mu_t], \tag{5.2.2}$$

so that the framework was able to cope with mass differences. Unfortunately, model (5.2.2) with source terms solely valued in $\mathcal{M}^+(\mathbb{R}^d)$ can not account for birth/death processes of individuals or cell divisions and transitions as we demonstrate in Example B.4 in Ref. [45]. To extend the framework to population dynamics we thus introduced an additional nonlinear growth/decay function $c : \mathbb{R}^d \times \mathcal{M}^+(\mathbb{R}^d) \to \mathbb{R}$, so that the model is of the form

$$\dot{\mu}_t = V[\mu_t] \oplus c(\cdot, \mu_t)\, \mu_t \oplus s[\mu_t]. \tag{5.2.3}$$

The notion $\oplus$ is applied to depict a summation of various effects, in this case transport, source and growth/decay processes. The rigorous meaning of (5.2.3) is based on the weak formulation of the problem, see Definition 5.2.1. The topic is constantly studied and we refer to [75] for a superposition principle, to [21] for a work on numerical schemes and to [48] for results on asymptotic stability and a Lyapunov function approach.

Definition 5.2.1. *We say that a continuous curve $\mu_\bullet : [0, T] \to (\mathcal{M}^+(\mathbb{R}^d), \|\cdot\|_{BL^*})$ is a* ***solution to*** (5.2.3) ***with initial condition*** $\mu_0 \in \mathcal{M}^+(\mathbb{R}^d)$ *if for all $f \in C_c^\infty(\mathbb{R}^d)$ and for all $t \in [0, T]$, it holds*

$$\begin{aligned} &\int_{\mathbb{R}^d} f(x)\,\mathrm{d}\mu_t(x) - \int_{\mathbb{R}^d} f(x)\,\mathrm{d}\mu_0(x) \\ &= \int_0^t \int_{\mathbb{R}^d \times \mathbb{R}^d} \nabla f(x) \cdot v \,\mathrm{d}V[\mu_r](x, v)\,\mathrm{d}r + \int_0^t \int_{\mathbb{R}^d} f(x)\, c(x, \mu_r)\,\mathrm{d}\mu_r(x)\,\mathrm{d}r \\ &\quad + \int_0^t \int_{\mathbb{R}^d} f(x)\,\mathrm{d}s[\mu_r](x)\,\mathrm{d}r. \end{aligned}$$

Remark 5.2.2. Although it may not immediately appear so, we can incoporate MDEs into the framework studied in Chapter 3 as in most cases the dynamics of MDEs simplifies to a transport equation in the spaces of measures. More specifically, we note that by Disintegration Theorem, see e.g. [53, Theorem 1.45], there exists a family of probability measures $\{\nu_{x,t}\}_{x\in\mathbb{R}^d, t\in[0,T]}$ such that

$$\int_0^t \int_{\mathbb{R}^d\times\mathbb{R}^d} \nabla f(x)\cdot v \,\mathrm{d}V[\mu_r](x,v)\,\mathrm{d}r = \int_0^t \int_{\mathbb{R}^d} \nabla f(x)\cdot \left[\int_{\mathbb{R}^d} v\,\mathrm{d}\nu_{x,r}(v)\right] \mathrm{d}x\,\mathrm{d}r$$

so that Definition 5.2.1 boils down to measure solutions for transport equation with velocity

$$\mathcal{V}(t,x) := \left[\int_{\mathbb{R}^d} v\,\mathrm{d}\nu_{x,t}(v)\right]. \tag{5.2.4}$$

From the MDE perspective, this representation provides no new insights as $\left[\int_{\mathbb{R}^d} v\,\mathrm{d}\nu_{x,t}(v)\right]$ cannot be computed explicitly in general. However, by a specific choice of b we can achieve a transport velocity in the spirit of (5.2.4), i.e. for model functions $\beta_i \in BL$ and $\nu_i \in \mathcal{M}^+$ define

$$b(\tau,x,\mu_t) = \begin{pmatrix} \int_{\mathbb{R}^d} \beta_1(\tau,x,v)\,\mathrm{d}[\nu_1(\tau,x,\mu_t)](v) \\ \vdots \\ \int_{\mathbb{R}^d} \beta_d(\tau,x,v)\,\mathrm{d}[\nu_d(\tau,x,\mu_t)](v) \end{pmatrix}.$$

So instead of prescribing a measure vector field V with corresponding (unknown) disintegrated measures $\nu_{x,r}$, we chose the opposite direction and prescribe the disintegrated measures which will then form a measure similar to the measure vector field. In this case, the ith component of b can be interpreted as a distribution of velocities in the ith dimension at point $x\in\mathbb{R}^d$ and time point $t\in[0,T]$. As usual, we are however not restricted to probability distributions. If we assume for all $i=1,...,d$

- $\beta_i \in BL([0,T]\times\mathbb{R}^d\times\mathbb{R}^d)$,
- $\nu_i \in BL([0,T]\times\mathbb{R}^d\times\mathcal{M}^+(\mathbb{R}^d);\mathcal{M}^+(\mathbb{R}^d))$,
- for all $R>0$ we have

$$\sup_{\substack{t\in[0,T]\\ \|\mu\|_{BL^*}\le R}} \|\nu_i(\tau,\cdot,\mu)\|_{BL^*} < \infty,$$

then it can be easily checked that b satisfies Assumptions 3.5.1.

5.3 A structured coagulation-fragmentation equation

In this section we show that a small adjustment of our theory enables to incorporate coagulation-fragmentation processes. Coagulation-fragmentation phenomena appear in various areas of science such as Physics [25, 43], Chemistry [169], Biology [10, 113] or Astronomy [132] and is still a very active field of research [59, 60, 100]. See [34] for a mathematical overview to the theory and numerous approaches.
In short, coagulation describes the process of mass (or size) accumulation of particles by their collision with other similar but smaller particles. In contrast, fragmentation characterises the break-up of a bigger particle into two or more smaller ones.

Usually, coagulation-fragmentation models are formulated with either a continuous or discrete structuring variable in $\mathbb{R}^d$. However, recently in [2] the authors translated the model formulation to the space of Radon measures under the flat norm and were thus able to unify both modelling approaches. Unfortunately, our setting - as presented in Section 3.5 - does not provide the correct model setup for formulating the coagulation-fragmentation model, so that the authors in [2] had to use the well-posedness theory developed in [3, 4]. To highlight where our model setup fails, we first introduce the notion of a weak solution to the coagulation-fragmentation model as it is presented in [2].
A map $\mu_\bullet \in C([0,T];(\mathcal{M}^+(\mathbb{R}^+),\|\cdot\|_{BL^*}))$ is called a **weak solution to the coagulation-fragmentation model** if for all $\varphi \in BL([0,T]\times\mathbb{R}^+)\cap C^1([0,T]\times\mathbb{R}^+)$ and for all $t \in [0,T]$ the following holds

$$\begin{aligned}
\int_{\mathbb{R}^+} \varphi(t,x)\,\mathrm{d}\mu_t(x) - \int_{\mathbb{R}^+} \varphi(0,x)\,\mathrm{d}\mu_0(x) &= \int_0^t \int_{\mathbb{R}^+} \partial_s\varphi(s,x)\,\mathrm{d}\mu_s(x)\,\mathrm{d}s \\
&+ \int_0^t \int_{\mathbb{R}^+} g(s,\mu_s)(x)\partial_x\varphi(s,x) - d(s,\mu_s)(x)\varphi(s,x)\,\mathrm{d}\mu_s(x)\,\mathrm{d}s \\
&+ \int_0^t \int_{\mathbb{R}^+} \int_{\mathbb{R}_+} \kappa(y,x)\left(\frac{1}{2}\varphi(s,x+y) - \varphi(s,x)\right)\mathrm{d}\mu_s(y)\,\mathrm{d}\mu_s(x)\,\mathrm{d}s \\
&+ \int_0^t \int_{\mathbb{R}^+} \left[\left(\int_0^x \varphi(s,y)\,\mathrm{d}[b(x)](y)\right)a(x) - a(x)\varphi(s,x)\right]\mathrm{d}\mu_s(x)\,\mathrm{d}s \\
&+ \int_0^t \int_{\mathbb{R}^+} \varphi(s,0)\beta(s,\mu_s)(x)\,\mathrm{d}\mu_s(x)\,\mathrm{d}s,
\end{aligned} \tag{5.3.1}$$

where

$$
\begin{aligned}
&g, d, \beta : [0,T] \times \mathcal{M}^+(\mathbb{R}^+) \to BL(\mathbb{R}^+), && a : \mathbb{R}^+ \to \mathbb{R}^+ \\
&b : \mathbb{R}^+ \to \mathcal{M}^+(\mathbb{R}^+). && \kappa : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+,
\end{aligned}
$$

As usual the functions g, d, β represent **growth, death and birth** functions and the symmetric coagulation kernel $\kappa(x, y)$ indicates the rate at which particles of size x and y coagulate. The function $a(x)$ denotes the **global fragmentation rate** of particles with size x and the measure $b(x)(A)$ supported in $[0, x]$ represents how many particles of size x fragment into a particle with size in the set A.

At first glance the notion of solutions fits almost perfectly to our framework. In view of Theorem 2.3.3 narrow convergence coincides with continuity in the space $(\mathcal{M}^+, \rho_F)$ and in view of Lemma 3.1.4 it doesn't matter if we consider the weak formulation for all $t \in [0, T]$ or for fixed $t = T$.
The main reason for the failure of our setup lies in the coagulation term and the collision-with-smaller-particles constraint. Note that the test function φ is evaluated at the point $x + y$ which can not be covered by any mutation kernel η in our weak formulation (3.5.2). To see where this stems from, we derive the weak formulation for the coagulation term by following [2]. In the continuous setting coagulation of particles with density u is usually described by a coagulation term of the form

$$
K(u)[x] = \int_0^x \frac{1}{2}\kappa(y, x-y)u(x-y)u(y)\,\mathrm{d}y - u(x)\int_0^\infty \kappa(y,x)u(y)\,\mathrm{d}y,
$$

Due to the collision-with-smaller-particles contraint κ depends both on y and on $x - y$. We multiply by a suitable smooth test function φ, integrate over the whole domain and apply Fubini Theorem

$$
\begin{aligned}
&\int_0^\infty \int_0^x \frac{1}{2}\kappa(y, x-y)u(x-y)u(y)\,\mathrm{d}y\,\varphi(x)\ \mathrm{d}x - \int_0^\infty u(x)\varphi(x)\int_0^\infty \kappa(y,x)u(y)\,\mathrm{d}y\,\mathrm{d}x \\
&= \int_0^\infty \int_y^\infty \frac{1}{2}\kappa(y, x-y)\varphi(x)u(x-y)\,\mathrm{d}x\,u(y)\,\mathrm{d}y - \int_0^\infty \int_0^\infty \kappa(y,x)u(y)\,\mathrm{d}y\,\varphi(x)u(x)\,\mathrm{d}x \\
&= \int_0^\infty \int_0^\infty \frac{1}{2}\kappa(y, x)\varphi(x+y)u(x)\,\mathrm{d}x\,u(y)\,\mathrm{d}y - \int_0^\infty \int_0^\infty \kappa(y,x)u(y)\,\mathrm{d}y\,\varphi(x)u(x)\,\mathrm{d}x.
\end{aligned}
\tag{5.3.2}
$$

We continue by assuming that u is actually the density of a measure μ with respect to the Lebesgue measure, i.e. $u(x) = D_\lambda\mu(x)$ and use basic properties of the Radon-

Nikodým derivative (see [44, F.22]) to rewrite (5.3.2) to

$$\int_0^\infty \int_0^\infty \frac{1}{2}\kappa(y,x)\varphi(x+y)\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) - \int_0^\infty \int_0^\infty \kappa(y,x)\,\mathrm{d}\mu(y)\,\varphi(x)\,\mathrm{d}\mu(x).$$

We notice that the argument $x-y$ in κ transformed to a shifted evaluation of the test function φ at the point $x+y$. As mentioned earlier, we can not choose η accordingly to account for (nonconstant) shifts in the evaluation of test functions.

Hence, a modification of our setup is necessary if we want to include coagulation-fragmentation models. In particular, we introduce an additional term which is similar to η but reflects the evaluation of test functions at a linear combination of variables. To this end, we first modify the weak formulation in the **linear case** (3.1.9) without state-independent influx N to

$$\begin{aligned}
&\int_{\mathbb{R}^d} \varphi(T,x)\,\mathrm{d}\mu_T(x) - \int_{\mathbb{R}^d} \varphi(0,x)\,\mathrm{d}\mu_0(x) = \int_0^T \int_{\mathbb{R}^d} \partial_t\varphi(t,x)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t \\
&\quad + \int_0^T \int_{\mathbb{R}^d} \big(\nabla_x\varphi(t,x)\cdot b(t,x) + \varphi(t,x)\,c(t,x)\big)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t \\
&\quad + \int_0^T \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \varphi(t,y)\,\mathrm{d}[\eta(t,x)]\,(y)\right)\mathrm{d}\mu_t(x)\,\mathrm{d}t \\
&\quad + \int_0^T \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \varphi(t,\alpha_1 x+\alpha_2 y)\,\mathrm{d}[K_{\alpha_1,\alpha_2}(t,x)](y)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t,
\end{aligned} \tag{5.3.3}$$

where $\alpha_1,\alpha_2\in\mathbb{R}$ and $K_{\alpha_1,\alpha_2}:[0,T]\times\mathbb{R}^d\to\mathcal{M}^+(\mathbb{R}^d)$ satisfies the same assumptions as η, see Assumptions 3.5.1. Of course, it is possible to include several terms of the form $K_{\alpha_1,\alpha_2}, K_{\beta_1,\beta_2}, \ldots$ in the formulation. The bounds of the theory will adapt accordingly as we show in the next paragraph. Note that choosing $\alpha_1=\alpha_2=1$ exactly covers the case which we need for the coagulation-fragmentation model.

Constructing a solution to the extended model:

We can solve (5.3.3) similar to (3.1.9) in Section 3.3 by considering the corresponding dual problem which in this case is given by

$$\begin{cases}
\partial_\tau\varphi_{\psi,t} + b\cdot\nabla_x\,\varphi_{\psi,t} + c\,\varphi_{\psi,t} + \int_{\mathbb{R}^d}\varphi_{\psi,t}(\tau,y)\,\mathrm{d}[\eta(\tau,x)]\,(y) & \\
\qquad + \int_{\mathbb{R}^d}\varphi_{\psi,t}(\tau,\alpha_1 x+\alpha_2 y)\,\mathrm{d}[K_{\alpha_1,\alpha_2}(\tau,x)](y) & = 0 \quad \text{in} \quad [0,t]\times\mathbb{R}^d, \\
\varphi_{\psi,t}(t,\cdot) & = \psi \quad \text{in} \quad \mathbb{R}^d,
\end{cases} \tag{5.3.4}$$

for some $t \in [0,T]$ and a fixed $\psi \in BL(\mathbb{R}^d)$. Via the method of characteristics, we get the following implicit integral equation for the solution $\varphi_{\psi,t}$

$$
\begin{aligned}
&\varphi_{\psi,t}(\tau,x) = \psi(X_b(t;\tau,x))e^{\int_\tau^t c(r,X_b(r;\tau,x))\,\mathrm{d}r}\\
&+\int_\tau^t \int_{\mathbb{R}^d} \varphi_{\psi,t}(s,y)\,\mathrm{d}[\eta(s,X_b(s;\tau,x))](y)e^{\int_\tau^s c(r,X_b(r;\tau,x))\,\mathrm{d}r}\,\mathrm{d}s\\
&+\int_\tau^t \int_{\mathbb{R}^d} \varphi_{\psi,t}(s,\alpha_1 X_b(s;\tau,x)+\alpha_2 y)\,\mathrm{d}[K_{\alpha_1,\alpha_2}(s,X_b(s;\tau,x))](y)e^{\int_\tau^s c(r,X_b(r;\tau,x))\,\mathrm{d}r}\,\mathrm{d}s,
\end{aligned}
$$

by defining

$$
C(t-\tau,x) = \int_{\mathbb{R}^d} \varphi_{\psi,t}(\tau,y)\,\mathrm{d}[\eta(\tau,x)](y) + \int_{\mathbb{R}^d} \varphi_{\psi,t}(\tau,\alpha_1 x+\alpha_2 y)\,\mathrm{d}[K_{\alpha_1,\alpha_2}(\tau,x)](y)
$$

in the proof of Proposition 3.3.1. As before, X_b denotes the flow generated by the vector field b which is defined as the unique solution of the ODE (3.2.1).
The subsequent proofs in Section 3.3 can be modified accordingly as the additional term basically behaves as the term involving η. In particular, after some calculations we can simply replace the term $\|\eta\|_\infty$ with $\|\eta\|_\infty + \|K_{\alpha_1,\alpha_2}\|_\infty$ in all estimates and get results which are comparable to 3.3.3 - 3.3.21.
Only for the nonnegativity proof of μ_t in Lemma 3.3.13 we remark that the introduced nonincreasing map $|\cdot|_- : \mathbb{R} \to \mathbb{R}^+$, $x \mapsto \max\{-x,0\}$ satisfies a triangle inequality, i.e. for all $a,b \in \mathbb{R}$

$$
|a+b|_- \leq |a|_- + |b|_-,
$$

so that the additional term involving K can be separated in (3.3.29) and treated independently which eventually leads again to an estimate of the form $\|\eta\|_\infty + \|K_{\alpha_1,\alpha_2}\|_\infty$.

As our modified linear model satisfies similar results to the ones obtained in Section 3.3, the generalisation of model (5.3.3) to the case with state-independent influx N in Section 3.4 and to the nonlinear model in Section 3.5 follows the very same lines as the proofs in the corresponding sections- apart from $\|\eta\|_\infty$ being replaced by $\|\eta\|_\infty + \|K_{\alpha_1,\alpha_2}\|_\infty$. For completeness, we provide the adjusted weak formulation in the nonlinear case, i.e. for $\varphi \in BL([0,T]\times\mathbb{R}^d)\cap C^1([0,T]\times\mathbb{R}^d)$

$$\begin{aligned}
&\int_{\mathbb{R}^d}\varphi(T,x)\,\mathrm{d}\mu_T(x)-\int_{\mathbb{R}^d}\varphi(0,x)\,\mathrm{d}\mu_0(x)=\int_0^T\int_{\mathbb{R}^d}\partial_t\varphi(t,x)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t\\
&\quad+\int_0^T\int_{\mathbb{R}^d}\big(\nabla_x\varphi(t,x)\cdot b(t,x,\mu_t)+\varphi(t,x)\,c(t,x,\mu_t)\big)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t\\
&\quad+\int_0^T\int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d}\varphi(t,y)\,\mathrm{d}[\eta(t,x,\mu_t)]\,(y)\right)\mathrm{d}\mu_t(x)\,\mathrm{d}t+\int_0^T\int_{\mathbb{R}^d}\varphi(t,x)\,\mathrm{d}[N(t,\mu_t)](x)\,\mathrm{d}t\\
&\quad+\int_0^T\int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d}\varphi(t,\alpha_1x+\alpha_2y)\,\mathrm{d}[K_{\alpha_1,\alpha_2}(t,x,\mu_t)](y)\right)\mathrm{d}\mu_t(x)\,\mathrm{d}t.
\end{aligned}\tag{5.3.5}$$

Remark 5.3.1. With the adjusted weak formulation (5.3.5) we are now able to transfer the coagulation-fragmention model from [2] to our setting. As domain we pick $\mathbb{R}^+$ and set

$$\begin{aligned}
b(t,x,\mu_t)=&g(t,\mu_t)(x),\\
c(t,x,\mu_t)=&-d(t,\mu_t)(x)-\int_{\mathbb{R}^+}\kappa(y,x)\,\mathrm{d}\mu_t(y)-a(x),\\
\eta(t,x,\mu_t)(\cdot)=&a(x)b(x)(\cdot)+\beta(t,\mu_t)(x)\delta_0(\cdot)\in\mathcal{M}^+(\mathbb{R}^+),\\
N(t,\mu_t)=&0\in\mathcal{M}^+(\mathbb{R}^+),\\
K_{1,1}(t,x,\mu_t)(\cdot)=&\frac{1}{2}\kappa(\cdot,x)\mu_t(\cdot)\in\mathcal{M}^+(\mathbb{R}^+),
\end{aligned}\tag{5.3.6}$$

where we abused the notation of the Radon-Nikodým derivative in $K_{1,1}$. More precisely, we assume that $1/2\,\kappa(\cdot,x)=D_{\mu_t}K_{1,1}(t,x,\mu_t)$ so that for all Borel sets $A\in\mathcal{B}(\mathbb{R}^+)$

$$K_{1,1}(t,x,\mu_t)=\int_A\frac{1}{2}\kappa(y,x)\,\mathrm{d}\mu_t(y).$$

Consequently, by basic properties of the Radon-Nikodým derivative (see [44, F.22]) we see

$$\int_{\mathbb{R}_+}\varphi(t,x+y)\,\mathrm{d}[K_{1,1}(t,x,\mu_t)](y)=\int_{\mathbb{R}_+}\varphi(t,x+y)\frac{1}{2}\kappa(y,x)\,\mathrm{d}\mu_t(y).\tag{5.3.7}$$

We plug the model functions (5.3.6) into (5.3.5) adjusted for the $\mathbb{R}^+$ setting. Furthermore, we use (5.3.7) and rearrange terms which leads to

$$
\begin{aligned}
&\int_{\mathbb{R}^+} \varphi(T,x)\,\mathrm{d}\mu_T(x) - \int_{\mathbb{R}^+} \varphi(0,x)\,\mathrm{d}\mu_0(x) = \int_0^T \int_{\mathbb{R}^+} \partial_t\varphi(t,x)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t \\
&\quad + \int_0^T \int_{\mathbb{R}^+} \partial_x\varphi(t,x) g(t,\mu_t)(x)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t \\
&\quad + \int_0^T \int_{\mathbb{R}^+} \varphi(t,x) \left(-d(t,\mu_t)(x) - \int_{\mathbb{R}_+} \kappa(y,x)\,\mathrm{d}\mu_t(y) - a(x)\right) \mathrm{d}\mu_t(x)\,\mathrm{d}t \\
&\quad + \int_0^T \int_{\mathbb{R}^+} \left(\int_{\mathbb{R}^+} \varphi(t,y)\,\mathrm{d}[a(x)b(x)(\cdot) + \beta(t,\mu_t)(x)\delta_0(\cdot)]\,(y)\right) \mathrm{d}\mu_t(x)\,\mathrm{d}t \\
&\quad + \int_0^T \int_{\mathbb{R}^+} \left(\int_{\mathbb{R}_+} \varphi(t,x+y)\,\mathrm{d}\left[\frac{1}{2}\kappa(\cdot,x)\mu_t(\cdot)\right](y)\right) \mathrm{d}\mu_t(x)\,\mathrm{d}t \\
&= \int_0^T \int_{\mathbb{R}^+} \partial_t\varphi(t,x) + \partial_x\varphi(t,x) g(t,\mu_t)(x) - d(t,\mu_t)(x)\varphi(t,x)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t \\
&\quad + \int_0^T \int_{\mathbb{R}^+} \int_{\mathbb{R}^+} \kappa(y,x) \left(\frac{1}{2}\varphi(t,x+y) - \varphi(t,x)\right) \mathrm{d}\mu_t(y)\,\mathrm{d}\mu_t(x)\,\mathrm{d}t \\
&\quad + \int_0^T \int_{\mathbb{R}^+} \left[\left(\int_{\mathbb{R}_+} \varphi(t,y)a(x)\,\mathrm{d}[b(x)]\,(y)\right) - \varphi(t,x)a(x)\right] \mathrm{d}\mu_t(x)\,\mathrm{d}t \\
&\quad + \int_0^T \int_{\mathbb{R}^+} \varphi(t,0)\beta(t,\mu_t)(x)\mu_t(x)\,\mathrm{d}t.
\end{aligned}
$$

This is exactly the weak model formulation (5.3.1) if we take into account that the measure $b(x)$ is only supported in $[0,x]$ by assumption.

5.4 A nonexpanding transport distance

We conclude this chapter by studying the transport distance introduced in the series of papers [64, 65]. The authors Fournier and Perthame used it to investigate the asymptotics of transport-type equations on a state space J, in particular to prove nonexpansiveness of the semigroup of solutions. As it turns out, in most cases the new distance is equivalent to a variant of the flat metric. In view of the concepts and methods developed in stochastic modelling, in particular the theory of concentrating Feller operators, the nonexpansiveness offers encouraging possibilities for the asymptotic analysis of measure solutions, see [99, 147, 148, 153].

In [65], the authors consider several transport-type models and for each equation introduce a model-specific distance to show nonexpansiveness. To this end, they require a **cost function** $\rho : J \times J \mapsto [0, \infty)$ which satisfies

- $\rho(x, x) = 0$,
- $\rho(x, y) = \rho(y, x) > 0$ for $x \neq y$.

Typically, the cost function will be of the form $\rho_a(x, y) = \min\{a, |x - y|\}$ for a suitably chosen $a > 0$. In accordance with optimal transport theory the authors then introduce a **transport cost** for probability measures related to the cost function ρ, i.e for $\mu_1, \mu_2 \in \mathcal{P}(J)$ they consider

$$\begin{cases} \mathcal{T}_\rho(\mu_1, \mu_2) = \inf_{\nu \in \mathcal{H}(\mu_1, \mu_2)} \int_J \int_J \rho(x, y)\, \nu(\mathrm{d}x, \mathrm{d}y), \\ \mathcal{H}(\mu_1, \mu_2) := \{\nu \in \mathcal{P}(J \times J) \text{ with marginals } \mu_1 \text{ and } \mu_2\} \end{cases} \tag{5.4.1}$$

Note that $\mathcal{T}_\rho$ defines a distance on $\mathcal{P}(J)$ if ρ is a metric on J.
We start with the following simple but important observation.

Lemma 5.4.1. *Let (J, d) be a metric space.*

i) For $a > 0$ consider the cost function $\rho_a(x, y) := \min\{a, d(x, y)\}$. Then ρ_a defines a distance on J as well.

ii) If (J, d) is a Polish metric space, then (J, ρ_a) is also a Polish metric space.

Proof. i) Positive definitness and symmetry directly follow from the respective properties of d. To prove the triangle inequality let $x, y, z \in J$ and we compute

$$\begin{aligned} \rho_a(x, y) &= \min\{a, d(x, y)\} \leq \min\{a, d(x, z) + d(z, y)\} \\ &\overset{(*)}{\leq} \min\{a, d(x, z)\} + \min\{a, d(z, y)\} = \rho_a(x, z) + \rho_a(z, y), \end{aligned}$$

where $(*)$ follows by a simple case distinction.
ii) The second statement follows immediately as for all $(x_n)_{n \in \mathbb{N}} \subset J$ and $x \in J$

$$d(x_n, x) \to 0 \qquad \Leftrightarrow \qquad \rho_a(x_n, x) \to 0 \qquad (n \to \infty) \tag{5.4.2}$$

so that d and ρ_a induce the same topology. $\square$

Remark 5.4.2. Let (J, d) be a Polish metric space and let ρ_a be as in Lemma 5.4.1. As (J, ρ_a) is also a Polish metric space by the previous lemma, the Kantorovich-Rubinstein duality (cf. [159, Remark 6.5]) implies that $\mathcal{T}_{\rho_a}$ is a Wasserstein distance for all $\mu_1, \mu_2 \in \mathcal{P}_1(J)$ with representation

$$\mathcal{T}_{\rho_a}(\mu_1, \mu_2) = \tilde{W}_1(\mu_1, \mu_2) := \sup_{|\psi|_{\mathbf{Lip}_{\rho_a}} \leq 1} \int_J \psi(x) \, \mathrm{d}(\mu_1 - \mu_2)(x). \tag{5.4.3}$$

Note, however, that the test functions ψ will be 1-Lipschitz continuous with respect to ρ_a and not with respect to d which we indicate by $|\cdot|_{\mathbf{Lip}_{\rho_a}}$. To emphasize the different set of test functions, we denote the resulting Wasserstein distance by $\tilde{W}_1$. Theorem 6.16 in [159] together with Lemma 5.4.1 implies that $(\mathcal{P}_1(J), \tilde{W}_1) = (\mathcal{P}_1(J), \mathcal{T}_{\rho_a})$ is also a Polish metric space.

We are now able to formulate our first equivalence result.

Proposition 5.4.3. *Let (J, d) be a Polish metric space and let $\tilde{\rho}_F$ denote the flat metric with underlying metric ρ_a, i.e.*

$$\tilde{\rho}_F(\mu_1, \mu_2) = \sup \left\{ \int_J \psi \, \mathrm{d}(\mu_1 - \mu_2) \mid \|\psi\|_\infty \leq 1, |\psi|_{\mathbf{Lip}_{\rho_a}} \leq 1 \right\}. \tag{5.4.4}$$

Then for all $\mu_1, \mu_2 \in \mathcal{P}_1(J)$

$$\tilde{\rho}_F(\mu_1, \mu_2) \leq \mathcal{T}_{\rho_a}(\mu_1, \mu_2) \leq \frac{1}{\max\{a, 1\}} \tilde{\rho}_F(\mu_1, \mu_2).$$

In particular, the distances $\mathcal{T}_{\rho_a}$ and $\tilde{\rho}_F$ are strongly equivalent on $\mathcal{P}_1(J)$.

Proof. Let $\mu_1, \mu_2 \in \mathcal{P}_1(J)$. According to Remark 5.4.2 we have the representation

$$\mathcal{T}_{\rho_a}(\mu_1, \mu_2) = \tilde{W}_1(\mu_1, \mu_2) = \sup_{|\psi|_{\mathbf{Lip}_{\rho_a}} \leq 1} \int_J \psi(x) \, \mathrm{d}(\mu_1 - \mu_2)(x).$$

This directly implies (5.4.4)

$$\tilde{\rho}_F(\mu_1, \mu_2) \leq \mathcal{T}_{\rho_a}(\mu_1, \mu_2).$$

For the other direction let ψ be a test function for $\mathcal{T}_{\rho_a}$ with $|\psi|_{\mathbf{Lip}_{\rho_a}} \leq 1$. We show that ψ also corresponds to a test function for $\tilde{\rho}_F$ by proving that the test functions are actually uniformly bounded. Indeed, without loss of generality the test functions for $\tilde{W}_1$ can be assumed to vanish uniformly at a prescribed point $x_0 \in J$. This is due to the fact that μ_1 and μ_2 are both probability measures, so that

$$\begin{aligned}\tilde{W}_1(\mu_1,\mu_2) &= \sup\left\{\int_J \psi \,\mathrm{d}(\mu_1-\mu_2) \mid |\psi|_{\mathbf{Lip}_{\rho_a}} \le 1\right\}\\ &= \sup\left\{\int_J \psi - \psi(x_0)\,\mathrm{d}(\mu_1-\mu_2) \mid |\psi|_{\mathbf{Lip}_{\rho_a}} \le 1\right\}\\ &= \sup\left\{\int_J \psi \,\mathrm{d}(\mu_1-\mu_2) \mid |\psi|_{\mathbf{Lip}_{\rho_a}} \le 1, \psi(x_0)=0\right\}.\end{aligned}$$

Now let ψ with $|\psi|_{\mathbf{Lip}_\rho} \le 1$ and $\psi(x_0)=0$. Then for all $x \in J$

$$|\psi(x)| \le |\psi(x)-\psi(x_0)| + |\psi(x_0)| \le \rho_a(x,y) + |\psi(x_0)| \le a,$$

so the test functions of $\tilde{W}_1$ are uniformly bounded. This leads to

$$\begin{aligned}\mathcal{T}_{\rho_a}(\mu_1,\mu_2) =& \tilde{W}_1(\mu_1,\mu_2) = \sup\left\{\int_J \psi \,\mathrm{d}(\mu_1-\mu_2) \mid |\psi|_{\mathbf{Lip}_{\rho_a}} \le 1, \psi(x_0)=0\right\}\\ =& \sup\left\{\int_J \psi \,\mathrm{d}(\mu_1-\mu_2) \mid \|\psi\|_\infty \le a, |\psi|_{\mathbf{Lip}_{\rho_a}} \le 1\right\}\\ \le& \frac{1}{\max\{a,1\}} \sup\left\{\int_J \psi \,\mathrm{d}(\mu_1-\mu_2) \mid \|\psi\|_\infty \le 1, |\psi|_{\mathbf{Lip}_{\rho_a}} \le 1\right\}\\ =& \frac{1}{\max\{a,1\}} \tilde{\rho}_F(\mu_1,\mu_2).\end{aligned}$$

□

Remark 5.4.4. One could ask also how the metric $\mathcal{T}_{\rho_a}$ relates to the “classical” flat metric (with respect to the underlying metric d). We first note that any test function which is 1-Lipschitz with respect to ρ_a is also 1-Lipschitz with respect to d since for all $x, y \in J$

$$|\psi(x) - \psi(y)| \le \rho_a(x,y) \le d(x,y).$$

Following the proof of Proposition 5.4.3 the test functions for $\tilde{W}_1$ can be assumed to vanish uniformly at an arbitrary point $x_0 \in J$ and are thus uniformly bounded.

Thus,

$$\begin{aligned}
\mathcal{T}_{\rho_a}(\mu_1,\mu_2) =& \tilde{W}_1(\mu_1,\mu_2) = \sup\left\{\int_J \psi \,\mathrm{d}(\mu_1-\mu_2) \mid |\psi|_{\mathbf{Lip}_{\rho_a}} \le 1, \psi(x_0)=0\right\}\\
=& \sup\left\{\int_J \psi \,\mathrm{d}(\mu_1-\mu_2) \mid |\psi|_{\mathbf{Lip}_{\rho_a}} \le 1, \|\psi\|_\infty \le a\right\}\\
\le& \sup\left\{\int_J \psi \,\mathrm{d}(\mu_1-\mu_2) \mid |\psi|_{\mathbf{Lip}_{d}} \le 1, \|\psi\|_\infty \le a\right\}\\
\le& \sup\left\{\int_J \psi \,\mathrm{d}(\mu_1-\mu_2) \mid \|\psi\|_{BL} \le \max\{a,1\}\right\}\\
=& \frac{1}{\max\{a,1\}}\rho_F(\mu_1,\mu_2).
\end{aligned}$$

In particular, the flat metric (with respect to the *basic* metric d) is stronger than $\mathcal{T}_{\rho_a}$. A reverse inequality is in general out of scope as Lipschitz continuity with respect to d does not imply Lipschitz continuity with respect to ρ_a, see also [136, 3.5]. However, if J is bounded, then the flat metric and the metric $\mathcal{T}_{\rho_a}$ are strongly equivalent on $\mathcal{P}_1(J)$.

To see this, we first note that for bounded J all metrics of the form $\rho_a(x,y) = \min\{a, d(x,y)\}$ are strongly equivalent. More precisely, for $0 < a \le b < \infty$ a short case distinction reveals

$$\rho_a(x,y) \le \rho_b(x,y) \le \frac{b}{a}\rho_a(x,y). \tag{5.4.5}$$

Now, we prove the equivalence of ρ_F and $\mathcal{T}_{\rho_a}$. Without loss of generality we assume that $a \le \operatorname{diam}(J)$ as we we could otherwise just consider some multiple of the diameter. Let ψ be a test function for the flat metric, i.e. $\|\psi\|_{BL} \le 1$. Then we compute for all $x, y \in J$ with (5.4.5)

$$|\psi(x)-\psi(y)| \le d(x,y) \le \min\{\operatorname{diam}(J), d(x,y)\} \le \frac{\operatorname{diam}(J)}{a}\rho_a(x,y),$$

so that the test functions are also Lipschitz continuous with respect to ρ_a. Hence we see

$$\begin{aligned}
\rho_F(\mu_1,\mu_2) =& \sup\left\{\int_J \psi \,\mathrm{d}(\mu_1-\mu_2) \mid \|\psi\|_\infty \le 1, |\psi|_{\mathbf{Lip}_d} \le 1\right\} \\
\le& \sup\left\{\int_J \psi \,\mathrm{d}(\mu_1-\mu_2) \mid \|\psi\|_\infty \le 1, |\psi|_{\mathbf{Lip}_{\rho_a}} \le \frac{\operatorname{diam}(J)}{a}\right\} \\
\le& \frac{a}{\operatorname{diam}(J)} \sup\left\{\int_J \psi \,\mathrm{d}(\mu_1-\mu_2) \mid |\psi|_{\mathbf{Lip}_{\rho_a}} \le 1\right\} \\
=& \frac{a}{\operatorname{diam}(J)} \mathcal{T}_{\rho_a}(\mu_1,\mu_2).
\end{aligned}$$

Proposition 5.4.5. *Let (S, d) be a Polish metric space. Then $\mathcal{T}_{\rho_a}$ metricises the narrow topology on $\mathcal{P}_1(J)$ (both with respect to ρ_a and d) and is thus (topologically) equivalent to ρ_F.*

Proof. According to Theorem 2.3.3, the flat metric ρ_F metricises the narrow topology on $\mathcal{M}^+(S)$, and thus also on $\mathcal{P}_1(S)$. On the other hand, Theorem 6.8 in Ref. [159] yields that the Wasserstein metric $\tilde{W}_1$, i.e. $\mathcal{T}_{\rho_a}$, induces the narrow topology on $\mathcal{P}_1(S)$. We remark that à priori the underlying test functions for narrow convergence differ in their notion of continuity for both metrics. More precisely, the test functions to check narrow continuity of ρ_F are continuous with respect to d, whereas the test functions for $\mathcal{T}_{\rho_a}$ are continuous for ρ_a. However, in view of (5.4.2) the topologies generated by ρ_a and d coincide, so there is no need to distiguish between the underlying metrics. In particular, $\mathcal{T}_{\rho_a}$ and ρ_F induces the same narrow topology. □

Example 5.4.6. The result of Proposition 5.4.3 can be applied to almost all examples in [65].

1. For the *general renewal equation* or *growth-fragmentation* on $J = [0, \infty)$ with transport cost $\rho(x, y) = \min\{a, d(x, y)\}$ the above results are directly applicable.

2. Similarly, this follows immediately for the *space and age structure* on $J = [0, \infty) \times \mathbb{R}^d$ with transport cost $\rho(x, z, y, r) = \min\{a, |x-y| + |z-r|\}$ as $|x-y| + |z-r|$ defines a metric on $\mathbb{R}^{d+1}$.

3. The same is true for the *multiple time renewal equation* on $J = \{(x_1, x_2) \in [0, \infty)^2 \mid x_2 > x_1\}$ with transport cost $\rho(x_1, x_2, \tilde{x}_1, \tilde{x}_2) = \min\{a, 2|x_1 - \tilde{x}_1| + |x_2 - \tilde{x}_2|\}$.

4. Consider the case of the *system of renewal equations* on $J = [0, \infty) \times T$ with $T = \{1, ..., I\}$ being the discrete Torus. The transport cost is given by $\rho(x, i, y, j) = \min\{a, d(x, y)\}\chi_{\{i=j\}} + a\chi_{\{i \neq j\}}$. This corresponds to the setting of J being I copies of the interval $[0, \infty)$ and the copies have uniform distance a from each other, so that we introduce

$$\tilde{J} = \bigotimes_{i=1}^{I} [0, \infty), \qquad \text{with} \qquad \tilde{d}(x^i, y^j) = \begin{cases} d(x^i, y^i), & i = j \\ a, & \text{else} \end{cases}$$

Then $\tilde{d}$ is clearly a metric $\tilde{J}$ and we can apply the above results.

Among the examples discussed in [65] the only transport cost which does not define a metric on $J = \mathbb{R}^d$ is in the case of *sexually structured populations* with cost $\rho(x, y) = |x - y|^p$ for some $p > 1$. ρ not being a metric implies that the Kantorovich-Rubinstein duality does not provide the useful representation (5.4.3) for $\mathcal{T}_\rho$. Consequently, the concept introduced by Fournier and Perthame is a true generalisation to our setting in this case.

6 Conclusion and outlook

The aim of this thesis was the study of first order nonlinear PDEs of the form

$$\begin{cases} \partial_t \mu_t + \nabla_x \cdot (b(t,x,\mu_t)\mu_t) & = c(t,x,\mu_t)\mu_t + \int_{\mathbb{R}^d} \eta(t,x,\mu_t)(y)\,\mathrm{d}\mu_t(y) + N(t,\mu_t), \\ \mu_0 & = \nu, \end{cases}$$

and how they can be appropriately extended to Polish metric spaces. To this end, first of all, key functional analytical properties of the space of Radon measures under the flat norm were established. Furthermore, the behaviour of the flat norm could be characterised more accurately by several results, such as the closed formula provided in Proposition 2.4.8 for the distance between a Dirac measure and a linear combination thereof. In most cases, however, the explicit calculation of the flat norm is difficult in the absence of a general analytical ground truth. This problem has been addressed by a recent approach using neural networks, with the main concepts briefly discussed here in this thesis.

In the second part of the thesis, the above PDE was examined on the Euclidean space $\mathbb{R}^d$ in a weak sense. To start with, a linear version of the model was considered that exhibited no state-independent influx N, which results in the associated differential operator being linear. Consequently, one could pass to the dual problem and derive an implicit integral representation for the dual solution $\varphi_{\psi,t}$ using the method of characteristics. The dual problem was then solved by a fixed point argument and its solution $\varphi_{\psi,t}$ was used to construct a solution to the primal problem. However, since the model functions are not autonomous, delicate regularity results for $\varphi_{\psi,t}$ were required for this step. In particular, one had to establish Fréchet differentiability of $\varphi_{\psi,t}$ in the parameter t using the Implicit Function Theorem in order to show that the constructed solution μ_t is indeed solving the weak formulation. Subsequently, using Duhamel's principle, a solution to the linear model with state-independent influx N was established, and uniqueness of solutions was proven. Finally, the non-linear problem could be solved through a reduction to a suitable linear problem.

The next chapter of the thesis dealt with the extension of the model to Polish metric

spaces and the corresponding well-posedness theory. For this purpose, a model formulation based on an implicit integral representation has been introduced to avoid the generally unavailable concept of PDEs. With this representation it was then possible to construct unique (generalised) solutions, first in the linear and then in the nonlinear case. It was also shown that the generalised modelling concept is a real extension of the PDE model, as both coincide in the Euclidean case.
In the final part of the thesis, the developed framework was then examined regarding applications. In addition to various sketches of modelling examples, it was also shown that measure differential equations can be represented. By a small modification of the weak formulation it is even possible to treat coagulation-fragmentation models.
In summary, this thesis has thoroughly explored the well-posedness of nonlinear models in measures, both for the first order nonlinear PDE version as well as the generalised approach on Polish metric spaces. The natural next step is now to investigate the asymptotics of measure solutions in more detail. In particular, the theory of concentrating Feller operators developed in Refs. [99, 147, 148] offers a promising ansatz for exploring the long-time behaviour of measures on Polish metric spaces. The foundation for this has already been laid by several results in this thesis. For example, uniform tightness of the solutions was established for all times under stronger modelling assumptions. Furthermore, it has been shown that the flat metric is topologically equivalent to the transport distance $\mathcal{T}_\rho$ introduced by Fournier and Perthame in Refs. [64, 65] to prove nonexpansiveness of conservative transport-type equations. Nonexpansiveness is a key property in Feller theory for proving the existence of an invariant measure, which in turn is the most promising candidate for an asymptotic limit of measure solutions. As an alternative approach, we mention the recently published paper Ref. [48] which deals with asymptotic stability of MDE solutions and Lyapunov functions. Possibly, these concepts can also be transfered to the framework considered in this thesis.

Related publications

Most of the results presented in Chapters 2 and 4 as well as Section 5.1 have already been presented in the following publications

C. Düll, P. Gwiazda, A. Marciniak-Czochra, J. Skrzeczkowski (2021) *Spaces of Measures and their Applications to Structured Population Models.* Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press,

C. Düll, P. Gwiazda, A. Marciniak-Czochra, J. Skrzeczkowski (2024) *Structured Population Models on Polish spaces: A unified Approach including Graphs, Riemannian Manifolds and Measure Spaces to describe Dynamics of Heterogeneous Populations.* Mathematical Models and Methods in Applied Sciences, 34 (1), p.109-143.

Section 2.5 summarises the content of the following preprint

H. Schmidt and C. Düll (2023) *Computing the distance between unbalanced distributions - the flat metric.* arXiv preprint:2308.01039, 2023.

Additionally, the following paper was produced during my doctoral studies and briefly addressed in Section 5.2

C. Düll, P. Gwiazda, A. Marciniak-Czochra, J. Skrzeczkowski (2023) *Measure differential equation with a nonlinear growth/decay term.* Nonlinear Analysis: Real World Applications, 73, 103917.

Bibliography

[1] A. S. Ackleh, R. M. Colombo, P. Goatin, S. Hille, and A. Muntean. Mathematical modeling with measures. *Math. Biosci. Eng.*, 17(3):2451–2452, 2020.

[2] A. S. Ackleh, R. Lyons, and N. Saintier. A structured coagulation-fragmentation equation in the space of Radon measures: unifying discrete and continuous models. *ESAIM Math. Model. Numer. Anal.*, 55(5):2473–2501, 2021.

[3] A. S. Ackleh and N. Saintier. Well-posedness of a system of transport and diffusion equations in space of measures. *J. Math. Anal. Appl.*, 492(1):124397, 28, 2020.

[4] A. S. Ackleh and N. Saintier. Diffusive limit to a selection-mutation equation with small mutation formulated on the space of measures. *Discrete Contin. Dyn. Syst. Ser. B*, 26(3):1469–1497, 2021.

[5] T. Alarcon, P. Getto, A. Marciniak-Czochra, and M. D. Vivanco. A model for stem cell population dynamics with regulated maturation delay. *Discrete Contin. Dyn. Syst.*, (Dynamical systems, differential equations and applications. 8th AIMS Conference. Suppl. Vol. I):32–43, 2011.

[6] J. D. Allan, M. Castillo Uzcanga, and K. Capps. *Stream Ecology, Structure and Function of Running Waters*. Springer Cham, third edition, 01 2021.

[7] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of bounded variation and free discontinuity problems*. Oxford Mathematical Monographs. The Clarendon Press, Oxford University Press, New York, 2000.

[8] K. H. Andersen. *Fish Ecology, Evolution, and Exploitation: A New Theoretical Synthesis*. Princeton University Press, 2019.

[9] C. Anil, J. Lucas, and R. Grosse. Sorting out Lipschitz function approximation. In K. Chaudhuri and R. Salakhutdinov, editors, *Proceedings of the 36th*

International Conference on Machine Learning, volume 97 of *Proceedings of Machine Learning Research*, pages 291–301. PMLR, 09–15 Jun 2019.

[10] O. Arino and R. Rudnicki. Phytoplankton dynamics. *Comptes Rendus Biologies*, 327(11):961–969, 2004.

[11] M. V. Barbarossa, M. Polner, and G. Röst. Stability switches induced by immune system boosting in an SIRS model with discrete and distributed delays. *SIAM J. Appl. Math.*, 77(3):905–923, 2017.

[12] M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396, 2003.

[13] N. Bellomo, J. A. Carillo, and E. Tadmor, editors. *Active particles, Vol. 3.* Modeling and Simulation in Science, Engineering and Technology. Birkhäuser/Springer, Cham, 2023. Advances in theory, models, and applications.

[14] N. Bellomo, P. Degond, and E. Tadmor, editors. *Active particles. Vol. 1. Advances in theory, models, and applications.* Modeling and Simulation in Science, Engineering and Technology. Birkhäuser Verlag/Springer-Verlag, Cham, 2017.

[15] N. Bellomo, P. Degond, and E. Tadmor, editors. *Active particles, Vol. 2.* Modeling and Simulation in Science, Engineering and Technology. Birkhäuser/Springer, Cham, 2019. Advances in theory, models, and applications.

[16] R. J. H. Beverton and S. J. Holt. On the dynamics of exploited fish populations. *Reviews in Fish Biology and Fisheries*, 4:259–260, 1993.

[17] A. Björck and C. Bowie. An iterative algorithm for computing the best estimate of an orthogonal matrix. *SIAM Journal on Numerical Analysis*, 8(2):358–364, 1971.

[18] V. I. Bogachev. *Measure theory. Vol. I.* Springer-Verlag, Berlin, 2007.

[19] J. Calder and M. Ettehad. Hamilton-Jacobi equations on graphs with applications to semi-supervised learning and data depth. *J. Mach. Learn. Res.*, 23(1), jan 2022.

[20] A. Calsina, O. Diekmann, and J. Z. Farkas. Structured populations with distributed recruitment: from PDE to delay formulation. *Math. Methods Appl. Sci.*, 39(18):5175–5191, 2016.

[21] F. Camilli, G. Cavagnari, R. De Maio, and B. Piccoli. Superposition principle and schemes for measure differential equations. *Kinet. Relat. Models*, 14(1):89–113, 2021.

[22] F. Camilli, A. Festa, and S. Tozza. A discrete Hughes model for pedestrian flow on graphs.

[23] J. A. Carrillo, R. M. Colombo, P. Gwiazda, and A. Ulikowska. Structured populations, cell growth and measure valued balance laws. *J. Differential Equations*, 252(4):3245–3277, 2012.

[24] G. Cavagnari, G. Savaré, and G. E. Sodini. Dissipative probability vector fields and generation of evolution semigroups in Wasserstein spaces. *Probab. Theory Related Fields*, 185(3-4):1087–1182, 2023.

[25] S. Chandrasekhar. Stochastic problems in physics and astronomy. *Rev. Mod. Phys.*, 15:1–89, Jan 1943.

[26] A. Chernodub and D. Nowicki. Norm-preserving orthogonal permutation linear unit activation functions (oplu), 2017.

[27] R. H. Chisholm, T. Lorenzi, A. Lorz, A. K. Larsen, L. N. d. Almeida, A. Escargueil, and J. Clairambault. Emergence of drug tolerance in cancer cell populations: An evolutionary outcome of selection, nongenetic instability, and stress-induced adaptation. *Cancer Research*, 75(6):930–939, 2015.

[28] L. Chizat, G. Peyré, B. Schmitzer, and F.-X. Vialard. Scaling algorithms for unbalanced optimal transport problems. *Math. Comp.*, 87(314):2563–2609, 2018.

[29] H. Cho, Y.-H. Kuo, and R. C. Rockne. Comparison of cell state models derived from single-cell RNA sequencing data: graph versus multi-dimensional space. *Mathematical Biosciences and Engineering*, 19(8):8505–8536, 2022.

[30] R. Colombo, M. Garavello, F. Marcellini, and E. Rossi. General renewal equations motivated by biology and epidemiology. *Journal of Differential Equations*, 354:133–169, 2023.

[31] E. D. Conway. Generalized solutions of linear differential equations with discontinuous coefficients and the uniqueness question for multidimensional quasilinear conservation laws. *J. Math. Anal. Appl.*, 18:238–251, 1967.

[32] J. M. Cushing. *An introduction to structured population dynamics*, volume 71 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1998.

[33] M. Cuturi. Sinkhorn Distances: Lightspeed Computation of Optimal Transport. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013.

[34] F. P. da Costa. Mathematical aspects of coagulation-fragmentation equations. In J.-P. Bourguignon, R. Jeltsch, A. A. Pinto, and M. Viana, editors, *Mathematics of Energy and Climate Change*, pages 83–162, Cham, 2015. Springer International Publishing.

[35] G. Da Prato and J. Zabczyk. *Stochastic equations in infinite dimensions*, volume 152 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, second edition, 2014.

[36] D.-P. Danciu, J. Hooli, A. Martin-Villalba, and A. Marciniak-Czochra. Mathematics of neural stem cells: Linking data and processes. *Cells & Development*, 174:203849, 2023.

[37] Z. a. Denkowski, S. a. Migórski, and N. S. Papageorgiou. *An introduction to nonlinear analysis: theory*. Kluwer Academic Publishers, Boston, MA, 2003.

[38] O. Diekmann, P. Getto, and M. Gyllenberg. Stability and bifurcation analysis of Volterra functional equations in the light of suns and stars. *SIAM J. Math. Anal.*, 39(4):1023–1069, 2007/08.

[39] O. Diekmann and M. Gyllenberg. *Abstract Delay Equations Inspired by Population Dynamics*, pages 187–200. Birkhäuser Basel, Basel, 2008.

[40] O. Diekmann, M. Gyllenberg, H. Huang, M. Kirkilionis, J. Metz, and H. Thieme. On the formulation and analysis of general deterministic structured population models II. nonlinear theory. *Journal of mathematical biology*, 43:157–89, 09 2001.

[41] O. Diekmann and J. A. P. Heesterbeek. *Mathematical epidemiology of infectious diseases*. Wiley Series in Mathematical and Computational Biology. John Wiley & Sons, Ltd., Chichester, 2000. Model building, analysis and interpretation.

[42] M. Doumic, A. Marciniak-Czochra, B. Perthame, and J. P. Zubelli. A structured population model of cell differentiation. *SIAM Journal on Applied Mathematics*, 71(6):1918–1940, 2011.

[43] W. Dreyer and F. Duderstadt. On the Becker/Döring theory of nucleation of liquid droplets in solids. *Journal of Statistical Physics*, 123:55–87, 2006.

[44] C. Düll, P. Gwiazda, A. Marciniak-Czochra, and J. Skrzeczkowski. *Spaces of Measures and their Applications to Structured Population Models*. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, 2021.

[45] C. Düll, P. Gwiazda, A. Marciniak-Czochra, and J. Skrzeczkowski. Measure differential equation with a nonlinear growth/decay term. *Nonlinear Anal. Real World Appl.*, 73:Paper No. 103917, 29, 2023.

[46] C. Düll, P. Gwiazda, A. Marciniak-Czochra, and J. Skrzeczkowski. Structured population models on Polish spaces: A unified approach including graphs, Riemannian manifolds and measure spaces to describe dynamics of heterogeneous populations. *Mathematical Models and Methods in Applied Sciences*, 34(01):109–143, 2024.

[47] J. Dyson, R. Villella-Bressan, and G. Webb. A nonlinear age and maturity structured model of population dynamics. I. Basic theory. *J. Math. Anal. Appl.*, 242(1):93–104, 2000.

[48] C. D'Apice, R. Manzo, B. Piccoli, and V. Vespri. Lyapunov stability for measure differential equations.

[49] H. Ennaji, Y. Quéau, and A. Elmoataz. Tug of War games and PDEs on graphs with applications in image and high dimensional data processing. *Scientific Reports*, 13, 2023.

[50] A. Esposito, F. S. Patacchini, and A. Schlichting. On a class of nonlocal continuity equations on graphs. *European Journal of Applied Mathematics*, 35(1):109–126, 2024.

[51] A. C. Esposito, F. S. Patacchini, A. Schlichting, and D. Slepčev. Nonlocal-interaction equation on graphs: Gradient flow structure and continuum limit. *Archive for Rational Mechanics and Analysis*, 240:699 – 760, 2021.

[52] L. C. Evans. *Partial differential equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, second edition, 2010.

[53] L. C. Evans and R. F. Gariepy. *Measure theory and fine properties of functions*. Textbooks in Mathematics. CRC Press, Boca Raton, FL, revised edition, 2015.

[54] J. Evers, S. C. Hille, and A. Muntean. Well-posedness and approximation of a measure-valued mass evolution problem with flux boundary conditions. *C. R. Math. Acad. Sci. Paris*, 352(1):51–54, 2014.

[55] J. H. M. Evers, S. C. Hille, and A. Muntean. Mild solutions to a measure-valued mass evolution problem with flux boundary conditions. *J. Differential Equations*, 259(3):1068–1097, 2015.

[56] J. H. M. Evers, S. C. Hille, and A. Muntean. Measure-valued mass evolution problems with flux boundary conditions and solution-dependent velocities. *SIAM J. Math. Anal.*, 48(3):1929–1953, 2016.

[57] J. Z. Farkas. Structured populations: the stabilizing effect of the inflow of newborns from an external source and the net growth rate. *Appl. Math. Comput.*, 199(2):547–558, 2008.

[58] J. Z. Farkas, P. Gwiazda, and A. Marciniak-Czochra. Asymptotic behaviour of a structured population model on a space of measures. *arXiv preprint arXiv:1902.06096*, 2019.

[59] M. Ferreira, E. Franco, and J.-J. Velazquez. On the self-similar behaviour of coagulation systems with injection, 06 2023.

[60] M. A. Ferreira, E. Franco, J. Lukkarinen, A. Nota, and J. J. L. Velázquez. Coagulation equations with source leading to anomalous self-similarity. *Journal of Physics A: Mathematical and Theoretical*, 56(48):485002, nov 2023.

[61] D. S. Fischer, A. K. Fiedler, E. M. Kernfeld, R. M. J. Genga, B.-P. Aimée, M. Bakhti, H. Lickert, J. Hasenauer, R. Maehr, and F. J. Theis. Inferring

population dynamics from single-cell RNA-sequencing time series data. *Nature Biotechnology*, 37(4):461–468, 2019.

[62] G. B. Folland. *Real analysis*. Pure and Applied Mathematics (New York). John Wiley & Sons, Inc., New York, 1984.

[63] G. B. Folland. *Real analysis*. Pure and Applied Mathematics (New York). John Wiley & Sons, Inc., New York, second edition, 1999. Modern techniques and their applications, A Wiley-Interscience Publication.

[64] N. Fournier and B. Perthame. Transport distances for PDEs: the coupling method. *EMS Surv. Math. Sci.*, 7(1):1–31, 2020.

[65] N. Fournier and B. Perthame. A nonexpanding transport distance for some structured equations. *SIAM J. Math. Anal.*, 53(6):6847–6872, 2021.

[66] M. D. Francesco, S. Fagioli, and E. Radici. Measure solutions, smoothing effect, and deterministic particle approximation for a conservation law with nonlocal flux. 2024.

[67] E. Franco, O. Diekmann, and M. Gyllenberg. Modelling physiologically structured populations: renewal equations and partial differential equations. *Journal of Evolution Equations*, 23, 06 2023.

[68] P. Gabriel, S. P. Garbett, V. Quaranta, D. R. Tyson, and G. F. Webb. The contribution of age structure to cell population responses to targeted therapeutics. *Journal of Theoretical Biology*, 311:19–27, 2012.

[69] A. Gayoso, P. Weiler, M. Lotfollahi, D. Klein, J. Hong, A. Streets, F. Theis, and N. Yosef. Deep generative modeling of transcriptional dynamics for RNA velocity analysis in single cells. *Nature Methods*, 21:50–59, 09 2023.

[70] C. Geiger and C. Kanzow. *Theorie und Numerik Restringierter Optimierungsaufgaben*. Masterclass. Springer Berlin, Heidelberg, 01 2002.

[71] P. Getto, M. Gyllenberg, Y. Nakata, and F. Scarabel. Stability analysis of a state-dependent delay differential equation for cell maturation: analytical and numerical methods. *Journal of Mathematical Biology*, 79, 07 2019.

[72] P. Getto and A. Marciniak-Czochra. *Mathematical Modelling as a Tool to Understand Cell Self-renewal and Differentiation*, pages 247–266. Springer New York, New York, NY, 2015.

[73] P. Gladbach, E. Kopfer, J. Maas, and L. Portinale. Homogenisation of one-dimensional discrete optimal transport. *Journal de Mathématiques Pures et Appliquées*, 139:204–234, 2020.

[74] D. R. Goethel and S. X. Cadrin. Revival and recent advancements in the spatial fishery models originally conceived by Sidney Holt and Ray Beverton. *ICES Journal of Marine Science*, 78(6):2298–2315, 05 2021.

[75] X. Gong and B. Piccoli. A measure model for the spread of viral infections with mutations. *Netw. Heterog. Media*, 17(3):427–442, 2022.

[76] K. Grauman and T. Darrell. Fast contour matching using approximate earth mover's distance. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, volume 1, 2004.

[77] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville. Improved training of Wasserstein GANs. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[78] P. Gwiazda, G. Jamróz, and A. Marciniak-Czochra. Models of discrete and continuous cell differentiation in the framework of transport equation. *SIAM Journal on Mathematical Analysis*, 44(2):1103–1133, 2012.

[79] P. Gwiazda, T. Lorenz, and A. Marciniak-Czochra. A nonlinear structured population model: Lipschitz continuity of measure-valued solutions with respect to model ingredients. *J. Differential Equations*, 248(11):2703–2735, 2010.

[80] P. Gwiazda and A. Marciniak-Czochra. Structured population equations in metric spaces. *J. Hyperbolic Differ. Equ.*, 7(4):733–773, 2010.

[81] P. Gwiazda, A. Marciniak-Czochra, and H. R. Thieme. Measures under the flat norm as ordered normed vector space. *Positivity*, 22(1):105–138, 2018.

[82] M. Gyllenberg and G. Webb. A nonlinear structured population model of tumor growth with quiescence. *Journal of mathematical biology*, 28(6):671–694, 1990.

[83] K. P. Hadeler. Neutral delay equations from and for population dynamics. In *The 8th Colloquium on the Qualitative Theory of Differential Equations*, volume 8 of *Proc. Colloq. Qual. Theory Differ. Equ.*, pages No. 11, 18. Electron. J. Qual. Theory Differ. Equ., Szeged, 2008.

[84] L. Haghverdi, F. A. Büttner, M.and Wolf, F. Buettner, and F. J. Theis. Diffusion pseudotime robustly reconstructs lineage branching. *Nature Methods*, 13:845–848, 2016.

[85] M. Hairer and J. C. Mattingly. Yet another look at Harris' ergodic theorem for Markov chains. In *Seminar on Stochastic Analysis, Random Fields and Applications VI*, volume 63 of *Progr. Probab.*, pages 109–117. Birkhäuser/Springer Basel AG, Basel, 2011.

[86] M. Hartvig, K. H. Andersen, and J. E. Beyer. Food web framework for size-structured populations. *J. Theoretical Biology*, 272(1):113–122, Mar 2011.

[87] S. C. Hille and D. T. H. Worm. Embedding of semigroups of Lipschitz maps into positive linear semigroups on ordered Banach spaces generated by measures. *Integral Equations and Operator Theory*, 63(3):351–371, 2009.

[88] M. Höcker, V. Berkhahn, A. Kneidl, A. Borrmann, and W. Klein. Graph-based approaches for simulating pedestrian dynamics in building models. In *eWork and eBusiness in Architecture, Engineering and Construction - Proceedings of the European Conference on Product and Process Modelling 2010*, pages 389–394. CRC Press, 2010.

[89] J. Jost. *Riemannian Geometry and Geometric Analysis.* Universitext. Springer, Cham, seventh edition, 2017.

[90] E. A. Karabut, E. N. Zhuravleva, and N. M. Zubarev. Application of transport equations for constructing exact solutions for the problem of motion of a fluid with a free boundary. *Journal of Fluid Mechanics*, 890:A13, 2020.

[91] M. Kimura and G. H. Weiss. The Stepping Stone Model of Population structure and the Decrease of genetic correlation with distance. *Genetics*, 49(4):561–576, 04 1964.

[92] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In Y. Bengio and Y. LeCun, editors, *3rd International Conference on Learning*

Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings, 2015.

[93] D. C. Kirouac, G. J. Madlambayan, M. Yu, E. A. Sykes, C. Ito, and P. W. Zandstra. Cell–cell interaction networks regulate blood stem and progenitor cell fate. *Molecular Systems Biology*, 5(1):293, 2009.

[94] A. Klenke. *Probability theory.* Universitext. Springer, Cham, third edition, 2020. A comprehensive course.

[95] A. Klimovskaia, D. Lopez-Paz, L. Bottou, and M. Nickel. Poincaré maps for analyzing complex hierarchies in single-cell data. *Nature Communications*, 11(1), 2020.

[96] L. B. Koralov and Y. G. Sinai. *Theory of probability and random processes.* Universitext. Springer, Berlin, second edition, 2007.

[97] M. Kwapisz. Bielecki's method, existence and uniqueness results for Volterra integral equations in L^p space. *J. Math. Anal. Appl.*, 154(2):403–416, 1991.

[98] G. La Manno, R. Soldatov, H. Hochgerner, A. Zeisel, V. Petukhov, M. Kastriti, P. Lönnerberg, A. Furlan, J. Fan, Z. Liu, D. van Bruggen, J. Guo, E. Sundström, G. Castelo-Branco, I. Adameyko, S. Linnarsson, and P. Kharchenko. Rna velocity in single cells. *nature*, 560:494–498, 8 2018.

[99] A. Lasota and T. Szarek. Lower bound technique in the theory of a stochastic differential equation. *J. Differential Equations*, 231(2):513–533, 2006.

[100] P. Laurençot. Stationary solutions to coagulation-fragmentation equations. *Annales de l'Institut Henri Poincaré C, Analyse non linéaire*, 36(7):1903–1939, 2019.

[101] J. M. Lee. *Introduction to Riemannian manifolds*, volume 176 of *Graduate Texts in Mathematics.* Springer, Cham, 1997.

[102] H. Ling and K. Okada. An efficient earth mover's distance algorithm for robust histogram comparison. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(5):840–853, 2007.

[103] Z. Long, Y. Lu, and B. Dong. Pde-net 2.0: Learning PDEs from data with a numeric-symbolic hybrid deep network. *Journal of Computational Physics*, 399:108925, 2019.

[104] Z. Long, Y. Lu, X. Ma, and B. Dong. PDE-net: Learning PDEs from data. In J. Dy and A. Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 3208–3216. PMLR, 10–15 Jul 2018.

[105] A. Marciniak-Czochra, T. Stiehl, A. Ho, W. Jäger, and W. Wagner. Modeling of asymmetric cell division in hematopoietic stem cells–regulation of self-renewal is essential for efficient repopulation. *Stem cells and development*, 18(3):377–386, 2009.

[106] V. Marot-Lassauzaie, B. J. Bouman, F. D. Donaghy, Y. Demerdash, M. A. G. Essers, and L. Haghverdi. Towards reliable quantification of cell state velocities. *PLOS Computational Biology*, 18(9):1–27, 09 2022.

[107] L. Marrec, I. Lamberti, and A.-F. Bitbol. Toward a universal model for spatially structured populations. *Phys. Rev. Lett.*, 127:218102, Nov 2021.

[108] M. Martcheva and S. S. Pilyugin. An epidemic model structured by host immunity. *J. Biological Systems*, 14(02):185–203, 2006.

[109] J. M. Mazón, M. Solera, and J. Toledo. Gradient flows in metric random walk spaces. *SeMA Journal*, 79:3 – 35, 2022.

[110] A. McKendrick. Application of mathematics to medical problems. *Proc. Edin. Math. Soc.*, 14:98–130, 1926.

[111] A. Moawad, A. Abbara, and A.-F. Bitbol. Evolution of cooperation in deme-structured populations on graphs. *Phys. Rev. E*, 109:024307, Feb 2024.

[112] M. Nicolau, A. J. Levine, and G. Carlsson. Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. *Proceedings of the National Academy of Sciences*, 108(17):7265–7270, 2011.

[113] H.-S. Niwa. School size statistics of fish. *Journal of Theoretical Biology*, 195(3):351–361, 1998.

[114] B. G. Pachpatte. *Inequalities for differential and integral equations*, volume 197 of *Mathematics in Science and Engineering.* Academic Press, Inc., San Diego, CA, 1998.

[115] R. K. Pedersen, M. Andersen, V. Skov, L. Kjær, H. C. Hasselbalch, J. T. Ottesen, and T. Stiehl. HSC Niche Dynamics in Regeneration, Pre-malignancy, and Cancer: Insights From Mathematical Modeling. *Stem Cells*, 41(3):260–270, 11 2022.

[116] B. Perthame. *Transport equations in biology.* Frontiers in Mathematics. Birkhäuser Verlag, Basel, 2007.

[117] G. Peyré and M. Cuturi. Computational optimal transport: With applications to data science. *Foundations and Trends in Machine Learning*, 11(5-6):355–607, 2019.

[118] G. Peyré, J. Fadili, and J. Rabin. Wasserstein active contours. In *2012 19th IEEE International Conference on Image Processing*, pages 2541–2544, 2012.

[119] B. Piccoli. Measure differential equations. *Arch. Ration. Mech. Anal.*, 233(3):1289–1317, 2019.

[120] B. Piccoli and F. Rossi. Generalized Wasserstein distance and its application to transport equations with source. *Arch. Ration. Mech. Anal.*, 211(1):335–358, 2014.

[121] B. Piccoli and F. Rossi. On properties of the generalized Wasserstein distance. *Arch. Ration. Mech. Anal.*, 222(3):1339–1365, 2016.

[122] B. Piccoli and F. Rossi. Measure-theoretic models for crowd dynamics. In *Crowd dynamics. Vol. 1*, Model. Simul. Sci. Eng. Technol., pages 137–165. Birkhäuser/Springer, Cham, 2018.

[123] B. Piccoli and F. Rossi. Measure dynamics with probability vector fields and sources. *Discrete Contin. Dyn. Syst.*, 39(11):6207–6230, 2019.

[124] R. Qu, X. Cheng, E. Sefik, J. III, B. Landa, F. Strino, S. Platt, J. Garritano, I. Odell, R. Coifman, R. Flavell, P. Myung, and Y. Kluger. Gene trajectory inference for single-cell data by optimal transport metrics. *Nature Biotechnology*, pages 1–11, 04 2024.

[125] R. Rabadan, Y. Mohamedi, U. Rubin, T. Chu, A. Alghalith, O. Elliott, L. Arnes, S. Cal, A. Obaya, A. Levine, and P. Cámara. Identification of relevant genetic alterations in cancer using topological data analysis. *Nature Communications*, 11, 07 2020.

[126] M. A. Rieffel. The Radon-Nikodym theorem for the Bochner integral. *Trans. Amer. Math. Soc.*, 131:466–487, 1968.

[127] F. Rihan. *Delay Differential Equations and Applications to Biology.* Springer Singapore, 08 2021.

[128] A. Rizvi, P. Camara, E. Kandror, T. Roberts, I. Schieren, T. Maniatis, and R. Rabadan. Single-cell topological RNA-Seq analysis reveals insights into cellular differentiation and development. *Nature Biotechnology*, 35, 05 2017.

[129] R. T. Rockafellar. *Conjugate duality and optimization.* Society for Industrial and Applied Mathematics, Philadelphia, Pa., 1974. Lectures given at the Johns Hopkins University, Baltimore, Md., June, 1973.

[130] F. Rossi, N. P. Duteil, N. Yakoby, and B. Piccoli. Control of reaction-diffusion equations on time-evolving manifolds. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, pages 1614–1619. IEEE, 2016.

[131] W. Rudin. *Functional Analysis.* International series in pure and applied mathematics. McGraw-Hill, 1991.

[132] V. S. Safronov. *Evolution of the protoplanetary cloud and formation of the earth and the planets.* Israel Program for Scientific Translations, 1972.

[133] G. Schiebinger, J. Shu, M. Tabaka, B. Cleary, V. Subramanian, A. Solomon, J. Gould, S. Liu, S. Lin, P. Berube, L. Lee, J. Chen, J. Brumbaugh, P. Rigollet, K. Hochedlinger, R. Jaenisch, A. Regev, and E. S. Lander. Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell*, 176(4):928–943.e22, 2019.

[134] D. K. Schlüter, I. Ramis-Conde, and M. A. J. Chaplain. Multi-scale modelling of the dynamics of cell colonies: insights into cell-adhesion forces and cancer invasion from in silico simulations. *Journal of The Royal Society Interface*, 12(103):20141080, 2015.

[135] H. Schmidt and C. Düll. Computing the distance between unbalanced distributions – the flat metric. *arXiv preprint arXiv:2308.01039*, 2023.

[136] D. R. Sherbert. Banach algebras of Lipschitz functions. *Pacific Journal of Mathematics*, 13(4):1387 – 1399, 1963.

[137] Y. N. Skiba. *Dynamics of Ideal Fluid on a Sphere*, pages 79–108. Springer International Publishing, Cham, 2017.

[138] J. Skrzeczkowski. Measure solutions to perturbed structured population models—differentiability with respect to perturbation parameter. *J. Differential Equations*, 268(8):4119–4182, 2020.

[139] A. A. Spector and W. L. Grayson. Stem cell fate decision making: Modeling approaches. *ACS biomaterials science & engineering*, 3 11:2702–2711, 2017.

[140] R. Stephany and C. Earls. Pde-learn: Using deep learning to discover partial differential equations from noisy, limited data. *Neural Networks*, 174:106242, 2024.

[141] T. Stiehl, N. Baran, A. D. Ho, and A. Marciniak-Czochra. Clonal selection and therapy resistance in acute leukaemias: mathematical modelling explains different proliferation patterns at diagnosis and relapse. *Journal of The Royal Society Interface*, 11(94):20140079, 2014.

[142] T. Stiehl, N. Baran, A. D. Ho, and A. Marciniak-Czochra. Cell Division Patterns in Acute Myeloid Leukemia Stem-like Cells Determine Clinical Course: A Model to Predict Patient Survival. *Cancer Research*, 75(6):940–949, 03 2015.

[143] T. Stiehl, C. Lutz, and A. Marciniak-Czochra. Emergence of heterogeneity in acute leukemias. *Biology Direct*, 11(51):1–19, 2016.

[144] T. Stiehl and A. Marciniak-Czochra. Stem cell self-renewal in regeneration and cancer: Insights from mathematical modeling. *Current Opinion in Systems Biology*, 5:112–120, 2017.

[145] T. Stiehl and A. K. Marciniak-Czochra. How to characterize stem cells? Contributions from mathematical modeling. *Current Stem Cell Reports*, 5:57–65, 2019.

[146] T. Stiehl, W. Wang, C. Lutz, and A. Marciniak-Czochra. Mathematical modeling provides evidence for niche competition in human aml and serves as a tool to improve risk stratification. *Cancer Research*, 80:canres.0283.2020, 07 2020.

[147] T. Szarek. Feller processes on nonlocally compact spaces. *Ann. Probab.*, 34(5):1849–1863, 2006.

[148] T. Szarek and D. T. H. Worm. Ergodic measures of Markov semigroups with the e-property. *Ergodic Theory Dynam. Systems*, 32(3):1117–1135, 2012.

[149] G. Teschl. *Ordinary differential equations and dynamical systems*, volume 140 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2012.

[150] H. R. Thieme. Discrete-time population dynamics on the state space of measures. *Math. Biosci. Eng.*, 17(2):1168–1217, 2020.

[151] H. R. Thieme. Persistent discrete-time dynamics on measures. In *Progress on difference equations and discrete dynamical systems*, volume 341 of *Springer Proc. Math. Stat.*, pages 59–100. Springer, Cham, 2020.

[152] H. R. Thieme. Discrete-time population dynamics of spatially distributed semelparous two-sex populations. *J. Math. Biol.*, 83(2):Paper No. 18, 40, 2021.

[153] H. R. Thieme. Discrete-time dynamics of structured populations via Feller kernels. *Discrete and Continuous Dynamical Systems - B*, 27(2):1091–1119, 2022.

[154] N. G. Trillos and D. Slepčev. Continuum limit of total variation on point clouds. *Archive for Rational Mechanics and Analysis*, 220:193 – 241, 2016.

[155] Y. Tsuzuku, I. Sato, and M. Sugiyama. Lipschitz-margin training: Scalable certification of perturbation invariance for deep neural networks. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018.

[156] A. Ulikowska. *Structured population models in metric spaces.* PhD thesis, University of Warsaw - Faculty of Mathematics, Informatics and Mechanics, 2013.

[157] C. Villa, P. K. Maini, A. P. Browning, A. L. Jenner, S. Hamis, and T. Cassidy. Reducing phenotype-structured PDE models of cancer evolution to systems of ODEs: a generalised moment dynamics approach. 2024.

[158] C. Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.

[159] C. Villani. *Optimal transport- old and new*, volume 338 of *Fundamental Principles of Mathematical Sciences*. Springer-Verlag, Berlin, 2009.

[160] H. Von Foerster. *Some remarks on changing populations*. The Kinetics of Cellular Proliferation. Shalton Press, New York, 1959.

[161] W. Walter. *Gewöhnliche Differentialgleichungen- Eine Einführung*. Springer, Berlin Heidelberg, seventh edition, 2000.

[162] L. White and G. Medley. Microparasite population dynamics and continuous immunity. *Proc.the R. Soc. Lond. Series B: Biological Sciences*, 265(1409):1977–1983, 1998.

[163] D. Wollny, S. Zhao, I. Everlien, X. Lun, J. Brunken, D. Brüne, F. Ziebell, I. Tabansky, W. Weichert, A. K. Marciniak-Czochra, and A. Martin-Villalba. Single-cell analysis uncovers clonal acinar cell heterogeneity in the adult pancreas. *Developmental cell*, 39 3:289–301, 2016.

[164] K. Yosida. *Functional analysis*. Classics in Mathematics. Springer-Verlag, Berlin, 1995. Reprint of the sixth (1980) edition.

[165] Y. Yuan and J. Bélair. Threshold dynamics in an SEIRS model with latency and temporary immunity. *J. Math. Biol.*, 69(4):875–904, 2014.

[166] L. Zappia, B. Phipson, and A. Oshlack. Splatter: simulation of single-cell RNA sequencing data. *Genome biology*, 18(1):1–15, 2017.

[167] S. Zhang, A. Afanassiev, L. Greenstreet, T. Matsumoto, and G. Schiebinger. Optimal transport analysis reveals trajectories in steady-state systems. *PLOS Computational Biology*, 17(12):1–29, 12 2021.

[168] Y. Zhang, T. Stiehl, and M. Tang. A continuous integral model for white blood cell production. *SIAM Journal on Applied Mathematics*, 82(6):2111–2130, 2022.

[169] R. M. Ziff and E. D. McGrady. Kinetics of polymer degradation. *Macromolecules*, 19:2513–2519, 1986.