Monge-Kantorovich Distances
and
Large-Time Asymptotics

Diplomarbeit
zur Erlangung des
akademischen Grades
”Magister der Naturwissenschaften”

Marcus Wunsch
Abstract. This diploma thesis shall be concerned with topics from optimal transportation theory, especially Monge-Kantorovich Distances, and their interrelation with the asymptotic behavior of both inviscid and viscous scalar conservation laws and with the stability of drift-diffusion-Poisson systems arising in semiconductor device modeling, on the real line.
## Contents

1. Acknowledgments 4  
2. Preface 4  

**Chapter 1. Introduction to the Optimal Transportation Problem.** 7  
1. **Existence of Optimal Transference Plans** 8  
2. **Topological Properties of Wasserstein Distances** 9  
3. **Convexity of Wasserstein Distances** 16  

**Chapter 2. W_p-Contractivity for Scalar Conservation Laws.** 21  
1. **Inviscid Scalar Conservation Laws** 21  
2. **Viscous Scalar Conservation Laws** 24  

**Chapter 3. W_2n-Stability for Drift-Diffusion-Poisson Systems.** 35  
1. **ddP-Systems with Linear Diffusion** 35  
2. **ddP-Systems with General Nonlinear Diffusion** 39  
3. **ddP-Systems for Plasmas** 40  

Bibliography 41
1. Acknowledgments

The author gratefully acknowledges financial support of P. A. Markowich’s Wittgenstein Award funded by the Austrian FWF.

2. Preface

At first sight, this thesis certainly seems somehow heterogeneous: We start with a general introduction to optimal transportation theory, then proceed with a contraction result for both inviscid and viscous scalar conservation laws on the real line, and finish with a stability result for a toy model of drift-diffusion-Poisson systems arising in semiconductor equations and plasmas, in $\mathbb{R}$ again. There is, however, a strong link between these three parts, established by the Wasserstein or Monge-Kantorovich distances.

It has become common to mention at this point that both the spelling and the attribution of Wasserstein distances can only vaguely be justified. Transliterating correctly from Russian, one should write “Vasershtain”, and there are many names of mathematicians who also discovered or rediscovered this distance, e.g. Gini, Höffding, Fréchet, Kantorovich and Rubinstein, and Tanaka. Despite this fact, the denomination of Wasserstein spaces has made its way into modern research literature, wherefore we shall mostly stick to it.

Optimal transportation theory is a branch of modern mathematics which, with some justification, can be called both ancient as well as a very young. It can be traced back to the dawn of the French Revolution, more exactly to the year of the publication of Gaspard Monge’s paper "Mémoire sur la théorie des déblais et des remblais", 1781. There he writes in plain prose:

"When one must transport soil from one location to another, the custom is to give the name clearing (déblais) to the volume of the soil that one must transport and the name filling (remblais) to the space that it must occupy after transfer.

Since the cost of transportation of one molecule is, all other things being equal, proportional to its weight and the interval that it must travel, and consequently the total cost of transportation being proportional to the sum of the products of the molecules each multiplied by the interval traversed: given the shape and position, the clearing and the filling, it is not the same for one molecule of the clearing to be moved to one or another spot of the filling. Rather, there is a certain distribution to be made of the molecules from the clearing to the filling, by which the sum of the products of molecules by intervals travelled will be the least possible, and the cost of the total transportation will be a minimum.” ([9],p.666, taken from [11])

In the forties of the past century, Leonid Kantorovich reformulated the optimal transportation problem by additionally requiring that no mass be split, and it was only in the eighties that mathematicians discovered the various connections of optimal transportation theory to their research fields. The link we are most interested in is that to partial differential equations.

Convergence results and also rates for many partial differential equations governing large-particle systems have been established by entropy-entropy dissipation
2. PREFACE

methods at the end of the nineties. So why bother trying new ways in order to get well-known results?

New approaches generally deepen the understanding of the subjects under investigation:
Various diffusion equations can, at least heuristically, be considered to be gradient flows on the space of probability measures, endowed with a manifold structure and a local metric whose arc length distance coincides with the quadratic Wasserstein distance. ([10])
Although convergence in Wasserstein sense may be weaker than convergence in entropy sense, this approach offers several advantages:
1) The Wasserstein distance is the natural distance associated to the gradient flow structure under examination.
2) The assumption of finite Wasserstein distance is much more general than the assumption of finite entropy.
3) Most importantly, by this method one can not only establish convergence to an equilibrium configuration, but directly compare two different solutions. Thus information about the short-time behavior of the gradient flow as well as its large-time asymptotics are gained. If there is contractivity, a fixed point argument (Banach Fixed Point Theorem) yields the existence and uniqueness of the stationary solution. ([6])

There are several mathematicians I owe my gratitude:
First of all, M. DiFrancesco for providing me with most of the mathematical background needed for my thesis, for his enduring patience answering my endless questions, and for leading me through the writing of this work.
G. Loeper was so kind as to give me some hints how the viscous case for scalar conservation laws could be treated.
J. Teichmann helped me with understanding the trickier parts of the topological properties of Wasserstein distances.
C. Villani himself read through all of this thesis, and I am very thankful for the enlightening discussions I had with his PhD-student F. Bolley on the contractivity of scalar conservation laws under $W_p$.
Last, but certainly not least, I want to thank P. A. Markowich for being my adviser.
CHAPTER 1

Introduction to the Optimal Transportation Problem.

Consider a pile of sand which has to be filled into a hole of equal size. Normalizing the mass of the pile to 1, we model the latter and the hole with probability measures \( \mu, \nu \), respectively, being supported in measurable spaces \( X, Y \). The effort of moving sand from the pile into the hole can be described by a measurable cost function \( c(x, y) \) defined on \( X \times Y \) being non-negative and possibly attaining infinite value. Naturally, the question arises how to realize the transportation at minimal cost.

Before devoting ourselves to this problem, we should specify what a way of transportation, or transference plan, actually is. We shall model transference plans by probability measures on \( X \times Y \). For a reasonable transference plan \( \pi \) on \( P(X, Y) \), one clearly has to require that all the mass taken from location \( x \) coincide with \( d\mu(x) \) and that all the mass transferred to \( y \) equal \( d\nu(y) \). In other words, for all \((\phi, \psi) \in C_b(X) \times C_b(Y)\), the equation

\[
\int_{X \times Y} [\phi(x) + \psi(y)]d\pi(x, y) = \int_X \phi(x)d\mu(x) + \int_Y \psi(y)d\nu(y)
\]

shall hold. We denote the set of all \( \pi \) satisfying this condition by \( \Pi(\mu, \nu) \) and say its elements have marginals \( \mu \) and \( \nu \) on \( X \) and \( Y \). We can now precisely formulate Kantorovich’s optimal transportation problem:

Minimize

\[
I[\pi] = \int_{X \times Y} c(x, y)d\pi(x, y), \quad \pi \in \Pi(\mu, \nu).
\]

For a given transference plan, \( I[\pi] \) is called the total transportation cost associated to \( \pi \). The optimal transportation cost between \( \mu \) and \( \nu \) is the quantity

\[
\mathcal{T}_c(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} I[\pi].
\]

The \( \pi \)'s attaining the infimum in \( \mathcal{T}_c \) - which do exist - for obvious reasons are called optimal transference plans.

There is also a very nice probabilistic interpretation of this problem, which is better known as optimal coupling of a pair of measures in probability theory: Given two probability measures \( \mu, \nu \), minimize the expectation

\[
I[U, V] = \mathbb{E}[c(U, V)],
\]

where \( U \) and \( V \) are random variables on \( X, Y \), with laws \( \mu, \nu \), respectively. This simply means \( \mu(A) = P(U^{-1}(A)) \) for all Borel-measurable \( A \subseteq X \). The optimal coupling is nothing but the pair \((U, V)\) minimizing the total transportation cost in \( \mathcal{T}_c \).
Next, we introduce the original mass transportation problem stated by G. Monge in 1781. Essentially, it is the same as Kantorovich’s problem, demanding additionally, however, that no mass be split: $d(x, y) = d(T(x), T(y)) = d(x)\delta(y = T(x))$, where $T : X \rightarrow Y$ is measurable. In terms of measurable subsets, the condition that $\pi_T$ belong to $\Pi(\mu, \nu)$ can be formulated as

for any measurable set $B \subseteq Y$, $\nu[B] = \mu[T^{-1}(B)]$.

Whenever $\pi_T$ satisfies this, we shall write $\nu = T_\# \mu$ and say that $\nu$ is the push-forward or (measure-theoretically) the image measure of $\mu$ by $T$. Recalling our definition of the law $\mu$ of a random variable $U$ on $(X, P)$, we observe that $\mu$ is nothing but $U_\# P$.

We are now in a position to state

Monge’s optimal transportation problem:

Minimize

$$I[T] = \int_X c(x, T(x))d\mu(x),$$

over all measurable maps $T$ such that $T_\# \mu = \nu$. At this point, two basic questions arise:

**Question 1:** Do minimizers of the Monge-Kantorovich problem exist, and if so, how can one characterize them?

**Question 2:** What information regarding $\mu, \nu$ can we derive from the knowledge of the optimal transportation cost $T_c(\mu, \nu)$?

### 1. Existence of Optimal Transference Plans

Since in this thesis, we shall not be mainly concerned with the first question, the following theorems asserting the existence of optimal transference plans are cited without proof.

**Theorem 1.1 (Knott-Smith optimality criterion).** Let $\mu, \nu$ be probability measures on $\mathbb{R}^n$ with finite second order moments. We consider the Monge-Kantorovich transportation problem associated to a quadratic cost function $c(x, y) = |x - y|^2$. Then $\pi \in \Pi(\mu, \nu)$ is optimal if and only if it is supported in the graph of the subdifferential of a convex lower semi-continuous function $\phi$, i.e. $\text{supp}(\pi) \subset \text{Graph}(\partial \phi)$.

**Proof.** See [13].

Let us remark that by the finiteness of the second order moments of $\mu, \nu$, we simply mean that

$$\int_{\mathbb{R}^n} \frac{|x|^2}{2} d\mu(x) + \int_{\mathbb{R}^n} \frac{|y|^2}{2} d\nu(y) < +\infty.$$  

Furthermore, the subdifferential $\partial \phi$ of a convex function $\phi$ is a set-valued application defined as

$$y \in \partial \phi(x) \iff \{\forall z \in \mathbb{R}^n, \phi(z) \geq \phi(x) + (y, z - x)\}.$$  

We shall also give the subsequent theorem by W. Gangbo and R. McCann on the existence of optimal transference plans for strictly convex costs.

**Theorem 1.2.** Let $c$ be a strictly convex, superlinear cost on $\mathbb{R}^n$, and let $\mu, \nu$ be probability measures on $\mathbb{R}^n$, such that the total transportation cost from $\mu$ to $\nu$ is not always infinite. Assume moreover that $\mu$ is absolutely continuous with respect to the Lebesgue measure. Then there exists a unique optimal transference plan for
2. Topological Properties of Wasserstein Distances

For the following considerations, the rather general setting of a Polish space $(X, d)$, that is, a separable and complete metric space, will suffice. We will denote by $P(X)$ the set of all Borel probability measures on $X$, defined on the Borel $\sigma$-algebra of $X$. Any Borel probability measure on a Polish space $X$ is regular and has $\sigma$-compact support. By definition, $\mu_k$ converges weakly to $\mu$ in $P(X)$ if for all bounded and continuous $Z$,

$$\lim_{k \to \infty} \int Z \, d\mu_k = \int Z \, d\mu.$$

This defines the weak topology on $P(X)$.

Moreover, Prokhorov’s theorem ensures that a subset $S$ of $P(X)$ is relatively weakly compact if and only if it is tight, that is, if for all $\varepsilon > 0$ there is a compact subset $K_\varepsilon$ of $X$ such that for all $\mu \in S$, $\mu(K_\varepsilon^c) \leq \varepsilon$ holds true. Now for the Polish space $(X, d)$, consider the convex cost functional $c(x, y) = d(x, y)^p$ for $p \geq 1$. We shall use the abbreviation $T_p(\mu, \nu) = T_p^\mu(\nu)$ (recall (1)) for the associated optimal transportation cost between two probability measures $\mu$ and $\nu$ on $X$. Henceforth, the set of all probability measures with finite moments of order $p$ will be denoted by $P_p(X)$.

**Theorem 2.1.** For all $p \in [1, +\infty)$, $W_p = T_p^{1/2}$ defines a metric on $P_p(X)$.

In order to verify the triangle inequality, we shall need the so-called Gluing Lemma:

**Lemma 2.2.** Let $\mu_1$, $\mu_2$, $\mu_3$ be three probability measures supported in Polish spaces $X_1$, $X_2$, $X_3$, respectively, and let $\pi_{12} \in \Pi(\mu_1, \mu_2)$, $\pi_{23} \in \Pi(\mu_2, \mu_3)$ be two transference plans. Then there is a probability measure $\pi \in P(X_1 \times X_2 \times X_3)$ with marginals $\pi_{12}$ on $X_1 \times X_2$ and $\pi_{23}$ on $X_2 \times X_3$.

**Proof.** When $X$ and $Y$ are Polish spaces, the concept of disintegration of measure establishes the possibility of writing any probability measure $\pi$ on $X \times Y$ with marginal $\mu$ on $X$ as

$$\pi = \int_X (\delta_x \otimes \pi_x) \, d\mu(x),$$

the Monge-Kantorovich transportation problem, and it has the form $\pi = (Id \times T)_\# \mu$, where $T$ is uniquely determined $\mu$-almost everywhere by the conditions $T_\# \mu = \nu$ and $T(x) = x - \nabla c^*(\nabla \phi(x))$ for some $c$-concave function $\phi$.

A clarification on terminology seems to be recommendable: By $c^*$ we denote the convex conjugate (or Legendre transform) of $c$: $\mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$, i.e.,

$$c^*(y) = \sup_{x \in \mathbb{R}^n} (xy - c(x)).$$

Moreover, a superlinear cost on $\mathbb{R}^n$ is a function $c$ satisfying

$$\lim_{|z| \to \infty} \frac{c(|z|)}{|z|} = +\infty$$

for all $z \in \mathbb{R}^n$.

**Proof.** See [13].
where the measurable application $x \mapsto \pi_x$, $X \to P(Y)$, is uniquely determined $d\mu$-almost everywhere. This of course means nothing but

$$\int_{X \times Y} u(x, y) = \int_X \left[ \int_Y u(x, y) d\pi_x(y) \right] d\mu(x)$$

for all $u \in C_b(X \times Y)$. Disintegrating both measures with respect to their common marginal $\mu_2$, we consequently write:

$$\pi_{12} = \int_{X_2} \pi_{12;2} \otimes \delta_{x_2} d\mu_2(x_2)$$

$$\pi_{23} = \int_{X_2} \delta_{x_2} \otimes \pi_{23;2} d\mu_2(x_2),$$

where $\pi_{12;2}$, $\pi_{23;2}$ are the corresponding measurable applications from $X_2$ into $P(X_1)$, $P(X_3)$, respectively. We now construct a new $\pi \in P(X_1 \times X_2 \times X_3)$ by

$$\pi = \int_{X_2} (\pi_{12;2} \otimes \delta_{x_2} \otimes \pi_{23;2}) d\mu_2(x_2).$$

This $\pi$ enjoys the required property of being an element of $\Pi(\pi_{12}, \pi_{23})$: for every $(\phi, \psi) \in C_b(X_1 \times X_2) \times C_b(X_1 \times X_2)$ we have

$$\int_{X_1 \times X_2 \times X_3} \left[ \phi(x_1, x_2) + \psi(x_2, x_3) \right] d\pi(x_1, x_2, x_3)$$

$$= \int_{X_3} d\pi_{23;2}(x_3) \left\{ \int_{X_2} \left[ \int_{X_1} \phi(x_1, x_2) d\pi_{12;2}(x_1) \right] d\mu_2(x_2) \right\}$$

$$+ \int_{X_3} \left\{ \int_{X_2} \left[ \int_{X_1} \psi(x_2, x_3) d\mu_2(x_2) \right] d\pi_{23;2}(x_3) \right\}$$

$$= \int_{X_2} \left[ \int_{X_1} \phi(x_1, x_2) d\pi_{12;2}(x_1) \right] d\mu_2(x_2)$$

$$+ \int_{X_3} \left[ \psi(x_2, x_3) d\mu_2(x_2) \right] d\pi_{23;2}(x_3)$$

$$= \int_{X_1 \times X_2} \phi(x_1, x_2) d\pi_{12}(x_1, x_2) + \int_{X_2 \times X_3} \psi(x_2, x_3) d\pi_{23}(x_2, x_3),$$

so the proof is complete. \qed

We are now in a position to check that $W_p$ actually defines a metric on $P_p(X)$.

**Proof.** From the definition it is clear that $W_p$ is symmetric, non-negative, and that $W_p(\mu, \mu) = 0$. On the other hand, let $\mu$, $\nu$ be two probability measures such that $W_p(\mu, \nu) = 0$. Let $\pi$ be an optimal transportation plan. Now $\text{supp}(d\pi(x, y))$


\[ f(y) = x, \quad \text{since} \quad 0 = W_p(\mu, \nu) = \mathcal{T}_p^2(\mu, \nu) = \left( \int_{X \times X} d(x, y)^p \, d\pi(x, y) \right)^{\frac{1}{p}} \]

\[ \Rightarrow \int_{X \times X} d(x, y)^p \, d\pi(x, y) = 0. \]

Thus, for all \( \phi \in C_b(X) \), \( \int \phi \, d\mu = \int \phi(x) \, d\pi(x, y) = \int \phi(y) \, d\pi(x, y) = \int \phi \, d\nu \), which implies \( \mu = \nu \).

In order to show the triangle inequality, let us consider \( \mu_1, \mu_2, \mu_3 \) in \( \mathcal{P}_p(X) \) and optimal transference plans \( \pi_{12} \) between \( \mu_1 \) and \( \mu_2 \) and \( \pi_{23} \) between \( \mu_2 \) and \( \mu_3 \). We choose \( X_i \) to be the support of \( \mu_i \), \( i=1,2,3 \). Let \( \pi \) be as in the gluing lemma, and let \( \pi_{13} \) be the marginal of \( \pi \) on \( X_1 \times X_3 \). Clearly, \( \pi_{13} \in \Pi(\mu_1, \mu_3) \). Hence we obtain:

\[ W_p(\mu_1, \mu_3) \leq \left( \int_{X_1 \times X_2} d(x_1, x_3)^p \, d\pi_{13}(x_1, x_3) \right)^{\frac{1}{p}} \]

\[ = \left( \int_{X_1 \times X_2 \times X_3} d(x_1, x_3)^p \, d\pi(x_1, x_2, x_3) \right)^{\frac{1}{p}} \]

\[ \leq \left( \int_{X_1 \times X_2 \times X_3} [d(x_1, x_2) + d(x_2, x_3)]^p \, d(x_1, x_2, x_3) \right)^{\frac{1}{p}} \]

\[ \leq \left( \int_{X_1 \times X_2 \times X_3} d(x_1, x_2)^p \, d\pi(x_1, x_2, x_3) \right)^{\frac{1}{p}} \]

\[ + \left( \int_{X_1 \times X_2 \times X_3} d(x_2, x_3)^p \, d\pi(x_1, x_2, x_3) \right)^{\frac{1}{p}} \]

\[ = \left( \int_{X_1 \times X_2} d(x_1, x_2)^p \, d\pi_{12}(x_1, x_2) \right)^{\frac{1}{p}} \]

\[ + \left( \int_{X_2 \times X_3} d(x_2, x_3)^p \, d\pi_{23}(x_2, x_3) \right)^{\frac{1}{p}} \]

\[ = W_p(\mu_1, \mu_2) + W_p(\mu_2, \mu_3), \]

so the triangle inequality holds true.

Let us now turn our attention towards the topological properties of Wasserstein distances. As a simple application of Hölder’s inequality, we notice that

\[ p_1 \geq p_2 \geq 1 \Rightarrow W_{p_1} \geq W_{p_2}, \]

because

\[ W_{p_2}(\mu, \nu)^{p_2} = \mathcal{T}_{p_2}(\mu, \nu) = \mathcal{T}_{p_2}(\mathcal{U}, \mathcal{V}) \]

\[ \geq \mathbf{E}(|\mathcal{U} - \mathcal{V}|)^{p_2} \geq \mathbf{E}(1)^{\frac{1}{2}} \mathbf{E}(|\mathcal{U} - \mathcal{V}|^{p_2})^{\frac{1}{2}} \]

\[ = W_{p_1}(\mu, \nu)^{\frac{p_2}{p_1}} \]

\[ \Rightarrow W_{p_2} \leq W_{p_1}. \]
12

1. Introduction to the Optimal Transportation Problem.

Here, $p^{-1} + q^{-1} = 1$, $p_1 = p_2 q > p_2$, and $\mathcal{U}, \mathcal{V}$ are random variables with law($\mathcal{U}$) = $\mu$, law($\mathcal{V}$) = $\nu$. Since $p, q$ were chosen arbitrarily, the statement holds for all $p_1, p_2$ with $p_1 \geq p_2 \geq 1$.

In the following theorem, we will encounter several equivalent statements of convergence in “Wasserstein sense”.

**Theorem 2.3 (Wasserstein Distances metrize Weak Convergence).** Let $p \in [1, +\infty)$, let $(\mu_k)_{k \in \mathbb{N}}$ be a sequence of probability measures in $P_p(X)$, and let $\mu \in P(X)$. Then, the following four statements are equivalent:

1. $W_p(\mu_k, \mu) \xrightarrow{k \to \infty} 0$,

2. $\mu_k \xrightarrow{k \to \infty} \mu$ weakly, and $(\mu_k)_{k \geq 1}$ satisfies the tightness condition: for some (and thus any) $x_0 \in X$,

   $\lim_{R \to +\infty} \limsup_{k \to +\infty} \int_{d(x_0, x) \geq R} d(x_0, x)^p \, d\mu_k(x) = 0$.

3. $\mu_k \xrightarrow{k \to \infty} \mu$ weakly, and there is convergence of order $p$: for some (and thus any) $x_0 \in X$,

   $\int d(x_0, x)^p \, d\mu_k(x) \xrightarrow{k \to \infty} \int d(x_0, x)^p \, d\mu(x)$.

4. Whenever a continuous function $\phi$ on $X$ satisfies the growth condition $|\phi| \leq C [1 + d(x_0, x)^p]$ for some $x_0 \in X$, $C \in \mathbb{R}$, then

   $\int \phi \, d\mu_k \xrightarrow{k \to \infty} \int \phi \, d\mu$.

**Proof.** We first check that [2], [3], and [4] are equivalent. Evidently, [4] implies [3]. Now assume that [2] is satisfied for some $x_0 \in X$ and arbitrarily choose a function $\phi$ obeying the growth condition in [4], and write

$\phi = \phi_R + \psi_R$

for $R > 1$, where $\phi_R(x) = \inf (\phi(x), C(1 + R^p))$, and $\psi_R(x) = \phi(x) - \phi_R(x)$ is pointwise bounded by $C d(x_0, x)^p 1_{d(x_0, x) \geq R}$. In order to show this, we distinguish between two cases: First of all, if $d(x_0, x) < R$, then

$\phi(x) < C(1 + R^p) \Rightarrow \phi_R \equiv \phi \Rightarrow \psi_R = 0$.

On the other hand, if $d(x_0, x) \geq R$, consequently we have

$\phi_R \equiv C(1 + R^p) \Rightarrow \psi_R \leq C(1 + d(x_0, x)^p) - C(1 + R^p) \leq C(d(x_0, x)^p)$. 
Hence we can estimate:

\[ \left| \int \phi \, d\mu_k - \int \phi \, d\mu \right| \leq \left| \int \phi_R \, d(\mu_k - \mu) \right| + C \int_{d(x_0, x) \geq R} d(x_0, x)^p \, d\mu_k(x) \]

\[ + C \int_{d(x_0, x) \geq R} d(x_0, x)^p \, d\mu(x). \]

Therefore,

\[ \limsup_{k \to \infty} \left| \int \phi \, d\mu_k - \int \phi \, d\mu \right| \leq C \limsup_{k \to \infty} \int_{d(x_0, x) \geq R} d(x_0, x)^p \left[ d\mu_k + d\mu \right](x). \]

Now the tightness condition in \( \mathbb{R}^d \) tells us that the right-hand side converges to 0 as \( R \) tends to infinity. This asserts that \( 2 \) implies \( 1 \).

In the subsequent step, let us show that \( 3 \) implies \( 2 \). Writing \( \bar{x} = \inf(\varphi, \psi) \), we deduce

\[ \int \left[ d(x_0, x) \wedge R \right]^p \, d\mu_k(x) \to \int \left[ d(x_0, x) \wedge R \right]^p \, d\mu(x); \]

on the other hand, the monotone convergence theorem gives

\[ \lim_{R \to \infty} \int \left[ d(x_0, x) \wedge R \right]^p \, d\mu(x) = \int d(x_0, x)^p \, d\mu(x). \]

Using the second condition in \( 3 \), we conclude that

\[ \lim_{R \to \infty} \lim_{k \to \infty} \int \left[ d(x_0, x)^p - (d(x_0, x) \wedge R)^p \right] \, d\mu_k(x) = 0. \]

Whenever \( d(x_0, x) \geq 2R \), we have \( d(x_0, x)^p - R^p \geq (1 - 2^{-p}) \, d(x_0, x)^p \). It follows that

\[ \lim_{R \to \infty} \limsup_{k \to \infty} \int_{d(x_0, x) \geq 2R} d(x_0, x)^p \, d\mu(x) = 0, \]

which is \( 2 \).

To complete the proof, it suffices to show that \( 1 \) \( \Leftrightarrow \) \( 3 \). The weak convergence of \( \mu_k \) to \( \mu \) implies

\[ \int d(x_0, x)^p \, d\mu(x) = \lim_{k \to \infty} \lim_{R \to \infty} \int \left[ d(x_0, x) \wedge R \right]^p \, d\mu_k(x) \]

\[ \leq \liminf_{k \to \infty} \int d(x_0, x)^p \, d\mu_k(x), \]

the equality being due to the monotone convergence theorem, and the inequality holding since \( d(x_0, x) \wedge R \) can be estimated from above by \( d(x_0, x) \), and since the limit, as \( k \to \infty \), is weak.

As a consequence, the convergence of the moment of order \( p \) in \( 3 \) is equivalent to

\[ \limsup_{k \to \infty} \int d(x_0, x)^p \, d\mu_k(x) \leq \int d(x_0, x)^p \, d\mu(x). \]

Next, we show that convergence in the Wasserstein sense implies \( 3 \). For this, we shall use the following elementary inequality:
For any \( \varepsilon > 0 \), there is a constant \( C_\varepsilon > 0 \) such that for all nonnegative real numbers \( a, b \),
\[(a + b)^p \leq (1 + \varepsilon) a^p + C_\varepsilon b^p.\]
Combination with the triangle inequality gives:
\[d(x_0, x)^p \leq [d(x_0, y) + d(x, y)]^p \leq (1 + \varepsilon) d(x_0, y)^p + C_\varepsilon d(x, y)^p.\]
If \( (\mu_k)_{k \in \mathbb{N}} \) is a sequence of probability measures in \( P_p(X) \) such that \( W_p(\mu_k, \mu) \)
tends to zero, and \( \pi_k \) is an optimal transference plan between \( \mu_k \) and \( \mu \), then let us integrate the last inequality with respect to \( \pi_k \) and use the marginal property. Then we get
\[
\int d(x_0, x)^p \ d\mu_k(x) \leq (1 + \varepsilon) \int d(x_0, y)^p \ d\mu(y) + C_\varepsilon \int d(x, y)^p \ d\pi_k(x, y).
\]
Because of the Wasserstein convergence of \( \mu_k \) to \( \mu \), the last summand vanishes as \( k \to \infty \). Therefore,
\[
\limsup_{k \to \infty} \int d(x_0, x)^p \ d\mu_k(x) \leq (1 + \varepsilon) \int d(x_0, x)^p \ d\mu(x).
\]
(6) now follows from letting \( \varepsilon \to 0 \).

It remains to show that (6) implies the weak convergence of \( \mu_k \) to \( \mu \), and that (6) implies (4). As a first step, let us prove these implications under the further assumption that \( d \) is bounded, say \( d \leq 1 \). So all the distances \( W_p \) are equivalent, and we just have to prove the theorem in the case \( p = 1 \). Then we use the Kantorovich-Rubinstein representation theorem, and (4) reduces to
\[
\sup_{\|\phi\|_{\text{Lip}} \leq 1} \int \phi \ d(\mu_k - \mu) \to 0.
\]
(5) Assume that \( W_1(\mu_k, \mu) \to 0 \). Let us prove that \( \mu_k \to \mu \) in the weak sense. From (5) we know that this is true if \( \phi \) is 1-Lipschitz (that is, \( \phi \)'s Lipschitz-norm is less than or equal to 1). Consequently, we can as well take any Lipschitz function \( \tilde{\phi} \) and replace it by \( \frac{\phi}{\|\phi\|_{\text{Lip}}} \) (as long as \( \phi \) is not equal to zero), and the argument still holds true. In a metric space, any bounded continuous function can be approximated from below and from above by a sequence of Lipschitz functions. More precisely, there is an isitone sequence \( (a_n)_{n \in \mathbb{N}} \) and an antitone sequence \( (b_n)_{n \in \mathbb{N}} \) of uniformly bounded Lipschitz functions, such that
\[
\lim_{n \to \infty} a_n = \phi = \lim_{n \to \infty} b_n.
\]
As a consequence,
\[
\limsup_{k \to \infty} \int \phi d\mu_k \leq \liminf_{n \to \infty} \limsup_{k \to \infty} \int b_n d\mu_k = \liminf_{n \to \infty} \int b_n d\mu = \int \phi d\mu,
\]
the inequality holding as \( \phi \) can be estimated pointwise from above by \( b_n \) by definition, the penultimate equality being due to the fact that all \( b_n \) are Lipschitz, and the last equality following from dominated convergence. Similarly, \( \liminf_{k \to \infty} \int \phi d\mu_k \geq \int \phi d\mu \), so weak convergence is actually established.

Conversely, consider a sequence \( (\mu_k)_{k \geq 1} \) converging weakly to \( \mu \), and prove (4). Let \( x_0 \) be an arbitrary element of \( X \), and denote by \( \text{Lip}_{1,x_0}^1 \) the set of all 1-Lipschitz functions with \( \phi(x_0) = 0 \). What we have to prove is that
\[
\sup_{\phi \in \text{Lip}_{1,x_0}^1} \int \phi d(\mu_k - \mu) \to 0.
\]
This will imply (1), because one can replace any \( \psi \in \text{Lip}_1(X) \setminus \text{Lip}_{1;x_0}(X) \) by \( \phi = \psi - \psi(x_0) \in \text{Lip}_{1;x_0} \) and in such a way extend the result.

As has already been mentioned above, Prokhorov’s theorem tells us that on Polish spaces, tightness and regularity of sequences of measures are equivalent. (Lower) regularity means that there exists a sequence of compact sets \( (K_n)_{n \geq 1} \) such that for all \( n \geq 1 \), we have \( \sup_{k \geq 1} \mu_k [K_n^c] \leq \frac{1}{n^2} \) and \( \mu [K_n^c] \leq \frac{1}{n} \). Without loss of generality, assume that \( x_0 \in K_1 \). Then, for each \( n \geq 1 \),

\[
\{ \phi_{1|K_n} | \phi \in \text{Lip}_{1;x_0}(X) \}
\]

is a subset of \( \text{Lip}_{1;x_0}(K_n) \), and by the Arzelà-Ascoli theorem, it is a compact subset of \( C_b(K_n) \), endowed with the norm of uniform convergence. Therefore, for any \( n \geq 1 \), we can extract a subsequence from any sequence in \( \text{Lip}_{1;x_0}(X) \) which converges uniformly on \( K_n \). By a diagonal argument, we arrive at the following statement: From any sequence \( (\phi_k)_{k \geq 1} \) in \( \text{Lip}_{1;x_0}(X) \) we can extract a subsequence which converges uniformly on each \( K_n \) to a function \( \phi_\infty \) defined on \( S = \bigcup K_n \), which is Lipschitz and bounded since the family \( (\phi_k) \) is uniformly bounded and uniformly Lipschitz.

We apply this statement to a family \( (\phi_k) \) satisfying

\[
\sup_{\phi \in \text{Lip}_{1;x_0}} \int \phi \, d(\mu_k - \mu) \leq \int \phi_k \, d(\mu_k - \mu) + \frac{1}{k}.
\]

So there is a subsequence, which we still denote by \( (\phi_k)_{k \in \mathbb{N}} \), that converges uniformly on each compact set \( K_n \) to a 1-Lipschitz function \( \phi_\infty \) on \( S \).

Next, we extend \( \phi_\infty \) to all of \( X \). To see that this is generally possible, take a 1-Lipschitz function \( F \) defined on a subset \( A_1 \) of a metric space \( A \) and write \( \tilde{F}(x) = \inf_{y \in A_1} (F(y) + d(x, y)) \), which again is a 1-Lipschitz function, however defined on all of \( A \). Our new function \( \phi_\infty \in \text{Lip}_{1;x_0} \) now enjoys the properties of continuity and boundedness (because the distance itself is bounded).

Let us show now that \( \int \phi_k \, d(\mu_k - \mu) \) tends to 0 as \( k \to \infty \). We compute

\[
\int \phi_k \, d(\mu_k - \mu) \leq \int_{K_n} (\phi_k - \phi_\infty) \, d(\mu_k - \mu) + \int_{K_n^c} (\phi_k - \phi_\infty) \, d(\mu_k - \mu) + \int_X \phi_\infty \, d(\mu_k - \mu).
\]

For any given \( n \), the first summand goes to 0 as \( k \to \infty \), because \( \phi_k \) converges uniformly to \( \phi_\infty \) on \( K_n \) as \( k \to \infty \). Since all \( \phi_k \) and \( \phi_\infty \) are uniformly bounded, the second term is bounded by \( C(\mu_k [K_n^c] + \mu [K_n^c]) \leq 2C/n \), for some constant \( C \), by the definition of \( (K_n) \), hence it converges uniformly to 0 as \( n \to \infty \). Finally, the third summand vanishes as \( k \to \infty \) for \( \mu_k \) converges weakly to \( \mu \). By letting first \( n \to \infty \), then \( k \to \infty \), we conclude the proof in the case of bounded distances \( d \).

To complete the proof, we shall show that we may as well take unbounded distances into account. For instance, take \( \bar{d} = \inf(d, 1) \) and let \( \bar{W}_p \) be the Wasserstein distance of order \( p \) constructed with \( \bar{d} \). It follows that \( W_p \geq \bar{W}_p \), so if convergence in \( W_p \) implies weak convergence, then so does \( W_p \)-convergence.

Conversely, assume that (3) is satisfied and that \( \bar{W}_p(\mu_k, \mu) \to 0 \). We shall use the
following inequality in order to prove that in this case, \( W_p(\mu_k, \mu) \to 0 \) as well:

\[
d(x, y) \leq d(x, y) \land R + 2d(x, x_0) \mathbb{1}_{d(x, x_0) \geq \frac{R}{2}} + 2d(x_0, y) \mathbb{1}_{d(x_0, y) \geq \frac{R}{2}}.
\]

We discern three cases: The first case, \( d(x, y) \leq \frac{R}{2} \), is trivial. Next, consider \( \frac{R}{2} \leq d(x, y) \leq R \), so our inequality becomes \( d(x, y) \leq R + C, C > 0 \), which again is evidently true. Finally, examine the case when \( d(x, y) > R \), and \( d(x_0, y) < \frac{R}{2} \) (which necessarily implies \( d(x, x_0) \geq \frac{R}{2} \)). By triangulation, this gives

\[
R < d(x, y) \leq d(x, x_0) + d(x_0, y) \leq d(x, x_0) + \frac{R}{2} \leq 2d(x, x_0) + R,
\]

which again asserts the inequality. Consequently, for some constant \( C(p) > 0 \), we have

\[
W_p(\mu_k, \mu) < R < d(x, y) \leq d(x, x_0) + d(x_0, y) \leq d(x, x_0) + \frac{R}{2} \leq 2d(x, x_0) + R;
\]

\[
(7)
\]

At last, we conclude that (1) is indeed satisfied, by letting first \( k \to \infty \), and then \( R \to \infty \).

\[\square\]

### 3. Convexity of Wasserstein Distances

For the Propositions to follow, we shall make use of the subsequent shorthand: Whenever \( \mathcal{U} \) and \( \mathcal{V} \) are two random variables with respective laws \( \mu \) and \( \nu \), we write

\[
\mathcal{T}_p(\mathcal{U}, \mathcal{V}) = \mathcal{T}_p(\mu, \nu)
\]

\[
W_p(\mathcal{U}, \mathcal{V}) = W_p(\mu, \nu)
\]

for the optimal transportation cost \( \mathcal{T}_p \) and the Wasserstein distance \( W_p \), which are associated to the cost function \( \|\cdot\|^p \). By definition, a random variable \( \mathcal{U} \) lies in \( L^p \) if \( E\|U\|^p < \infty \).
Proposition 3.1 (Behavior of the Wasserstein distance under rescaled convolution). Let $X$ be a normed space, $p \geq 1$. Whenever the random variables $U_1, U_2, V_1, V_2$ (valued in $X$) lie in $L^p$, $U_1$ is independent of $U_2$, $V_1$ is independent of $V_2$, and $\alpha_1, \alpha_2, \in \mathbb{R}$, then
\[ T_p(\alpha_1 U_1 + \alpha_2 U_2, \alpha_1 V_1 + \alpha_2 V_2) \leq 2^{p-1} [|\alpha_1|^p T_p(U_1, V_1) + |\alpha_2|^p T_p(U_2, V_2)]. \]

Proof. In this context, it is recommendable to work in the setting of random variables (cf. Chapter II (b)), and to write
\[ E[|\alpha_1 U_1 + \alpha_2 U_2| - (\alpha_1 V_1 + \alpha_2 V_2)]^p = E[\alpha_1 (U_1 - V_1) + \alpha_2 (U_2 - V_2)]^p. \]
Then we use the inequality $(a + b)^p \leq 2^{p-1}(a^p + b^p)$. In order to see that it holds, apply H"older’s inequality to the sum $a \cdot 1 + b \cdot 1$:
\[ a \cdot 1 + b \cdot 1 \leq (a^p + b^p)^{\frac{1}{p}}(1 + 1)^{\frac{1}{1-p}} \Rightarrow (a + b)^p \leq 2^{p-1}(a^p + b^p). \]
Now we can calculate that the last expectation is bounded from above by
\[ 2^{p-1}[|\alpha_1|^p E[U_1 - V_1]^p + |\alpha_2|^p E[U_2 - V_2]^p]. \]
In order to conclude the proof, we take the optimal couplings of $\alpha_1 U_1 + \alpha_2 U_2$ and $\alpha_1 V_1 + \alpha_2 V_2$ on the right hand side first, and then on the left hand side.

For the following Proposition, we will use the notation $U \perp V$ for saying that the random variables $U$, $V$ are independent.

Proposition 3.2 (Subadditivity of $T_2$ under rescaled convolution). Let $U_1 \perp U_2, V_1 \perp V_2$ be $L^2$-random variables with values in a Hilbert space $X$, such that either $EU_1 = EV_1$ or $EU_2 = EV_2$. Let $\alpha_1, \alpha_2 \in \mathbb{R}$. Then
\[ T_2(\alpha_1 U_1 + \alpha_2 U_2, \alpha_1 V_1 + \alpha_2 V_2) \leq \alpha_1^2 T_2(U_1, V_1) + \alpha_2^2 T_2(U_2, V_2). \]

Proof. Without loss of generality, the coupling $(U_1, V_1)$ may be chosen independent from $(U_2, V_2)$, since in the definition of the left-hand side of (10), only the laws of $\alpha_1 U_1 + \alpha_2 U_2$ and $\alpha_1 V_1 + \alpha_2 V_2$ matter. Expanding the square expectation, we get
\[ E[|\alpha_1 (U_1 - V_1) + \alpha_2 (U_2 - V_2)|^2] = \alpha_1^2 E[U_1 - V_1]^2 + \alpha_2^2 E[U_2 - V_2]^2 + 2\alpha_1 \alpha_2 E(U_1 - V_1, U_2 - V_2). \]
By independence, the last term equals $2\alpha_1 \alpha_2 (EU_1 - EV_1, EU_2 - EV_2)$, so it vanishes because of our above assumption. Then pass to the infimum as before.

Remark 3.3. The reason why we used the expression "rescaled convolution" in the previous two propositions will become evident from the following:

If $U, V : (\Omega, P) \to X$ are independent random variables and $\mu = \text{law}(U)$ and $\nu = \text{law}(V)$, then it is a well-known fact from probability theory that $\text{law}(X + Y) = \mu * \nu$, where * means convolution. Now observe that $\text{law}(\alpha U) = m_{\alpha \mu}$, where $m_\lambda$ stands for multiplication with the scalar $\lambda$:
\[ (m_{\alpha \mu})_1(B) = \mu_1(m_{\alpha}^{-1}(B)) = P(U^{-1}(m_{\alpha}^{-1}(B))) = P((\alpha X)^{-1}(B)) \]
for all Borel sets $B \in X$. Moreover,
\[ \text{law}(\alpha U + (1 - \alpha)V) = (m_{\alpha \mu}) * (m_{(1-\alpha)\nu}). \]
Finally, if $X = \mathbb{R}^n$, and $\mu$ has a density function $f$, then
\[ d(m_{\alpha \mu}) = d\mu \left[ m_{\mu}^{-1}(x) \right] = f \left( \frac{x}{\alpha} \right) d \left( \frac{x}{\alpha} \right) = \frac{1}{|\alpha|^n} f \left( \frac{x}{\alpha} \right) dx. \]
We recall that defining
\[ W_p(\mu, \nu)^p = W_{d\nu}(\mu, \nu)^p = \inf_{T_{\mu=\nu}} I[T], \]
we actually gain a distance that metrizes weak convergence on the space of probability measures on a Polish space \( X \) with finite \( p \)-th moment,
\[ P_p(X) = \{ \mu \in P(X) : \int_X |x|^p \, d\mu(x) < \infty \}. \]

Wasserstein distances of order \( p \) have recently been applied to one-dimensional diffusion equations, with cost function \( c(x, y) = |x-y|^p \). In this setting, Wasserstein distances have a delightful representation. For this purpose, let us concentrate on density functions \( \rho, \sigma \) rather than their associated probability measures \( \mu, \nu \). Then, let us define the corresponding distribution functions
\[ F(y) = \int_{-\infty}^{y} \rho(x) \, dx \]
\[ G(y) = \int_{-\infty}^{y} \sigma(x) \, dx, \]
as well as their pseudo-inverses \( F^{-1}, G^{-1} : (0, 1) \to \mathbb{R}, \)
\[ F^{-1}(\eta) = \inf_{\omega \in \mathbb{R}} \{ F(\omega) > \eta \} \]
\[ G^{-1}(\eta) = \inf_{\omega \in \mathbb{R}} \{ G(\omega) > \eta \}. \]

As can easily be proven, the optimal map \( T^* \) among all \( T_{\mu=\sigma} \) can be expressed as
\[ T^* = G^{-1} \circ F. \]

We would like to stress the fact that this \( T^* \) is the only map transporting \( \rho \) to \( \sigma \): We compute
\[ \int_{-\infty}^{+\infty} \rho(x) \varphi(x) \, dx = \int_{-\infty}^{+\infty} \sigma(T(x)) \varphi(T(x)) |T'(x)| \, dx \]
\[ \Rightarrow F(y) = \int_{-\infty}^{y} \rho(x) \, dx \]
\[ = \int_{-\infty}^{T(y)} \sigma(T(x)) |T'(x)| \, dx = G(T(y)), \]
where \( \varphi(x) \equiv 1_{(-\infty, x]} \). Hence, by slight abuse of notation, we have

\[
W_p(\rho, \sigma)^p = \int_{-\infty}^{\infty} |x - T^*(x)|^p \rho(x) \, dx
\]

\[
= \int_{-\infty}^{\infty} |x - G^{-1} \circ F(x)|^p \rho(x) \, dx
\]

\[
= \int_0^1 |F^{-1}(\eta) - G^{-1}(\eta)|^p \, d\eta
\]

by the change of variables \( F(x) = \eta, \rho(x) \, dx = d\eta \).

Therefore, the Wasserstein distance of order \( p \) between two density functions on the real line is nothing but the \( L^p((0,1)) \) - norm of their pseudo-inverses.
CHAPTER 2

$W_p$-Contractivity for Scalar Conservation Laws.

As a general reference for scalar conservation laws, we suggest e.g. [8], or [12]. Given a locally Lipschitz, real-valued function $f$ on $\mathbb{R}$, we shall consider the inviscid scalar conservation law

$$u_t + f(u)_x = 0, \quad t > 0, \quad x \in \mathbb{R}$$

$$u(0, x) = u^0(x),$$

(8)

as well as the viscous scalar conservation law

$$u_t + f(u)_x = u_{xx}, \quad t > 0, \quad x \in \mathbb{R}$$

$$u(0, x) = u^0(x).$$

(9)

We will be interested in the asymptotic behavior of both classical as well as entropy solutions and weak solutions of (8) and (9), respectively.

The proofs of the results, theorems, and propositions to follow can be found in [3].

1. Inviscid Scalar Conservation Laws

We will require the initial conditions to lie in a special subset $\mathcal{U} \subseteq L^\infty(\mathbb{R})$.

**Definition 1.1.** A function $v : \mathbb{R} \rightarrow \mathbb{R}$ will be said to lie in $\mathcal{U}$ if it is non-decreasing, right-continuous, and has the limits $0$ and $1$ at $-\infty$ and $+\infty$, respectively.

In other words, $\mathcal{U}$ is the set of all probability distribution functions on the real line. It is an interesting fact that this set is preserved by both conservation laws:

**Theorem 1.2.** Given the initial datum $u^0 \in \mathcal{U}$, the unique entropy solution $u$ of (8) in $L^\infty([0, +\infty) \times \mathbb{R}) \cap C([0, +\infty), L^1_{loc}(\mathbb{R}))$ is a.e. equal to an element of $\mathcal{U}$ to $\mathcal{U}$ for all $t \geq 0$.

**Proof.** See [12].

The map

$$\delta_x : v \mapsto v_x$$

(10)

$$v_x([-\infty, x)) = v(x)$$

is injective. $\mathcal{P}(\mathbb{R})$ stands for the sets of probability measures on the real line. Moreover, let us introduce the sets $\mathcal{U}_p$ and $\mathcal{P}_p$, for $p \geq 1$:

**Definition 1.3.** A function $v$ in $\mathcal{U}$ belongs to $\mathcal{U}_p$ if its distributional derivative $v_x$ satisfies $\int_{-\infty}^{+\infty} |x|^p \, dv_x(x) < +\infty$.

By $\mathcal{P}_p$ we denote the set of all probability measures having finite moment of order $p$. 


Again, the map
\[
U_p \to P_p(\mathbb{R})
\]
\[
\delta_x : v \mapsto v_x
\]
\[
v_x([-\infty, x]) = v(x)
\]
is one-to-one, so we can define a distance \(d_p\) on \(U_p \times U_p\) by
\[
d_p(v, \tilde{v}) = W_p(v_x, \tilde{v}_x).
\]
In the case of classical solutions of (8), the Wasserstein distance between two solutions is actually conserved. As a consequence, we have
\[
d_p(v(t, \cdot), \tilde{v}(t, \cdot)) = d_p(v(x), \tilde{v}(x)),
\]
for \(v, \tilde{v} \in U_p\).

The proof of this statement is surprisingly simple and shall be given here.

Consider the case of two classical solutions \(v, \tilde{v}\) such that \(v, \tilde{v} \in U\) and are increasing for all \(t \geq 0\). The map \(v^0\) is strictly increasing from 0 to 1, so it has a true inverse \(X(0, \cdot)\) defined on \((0, 1)\) by
\[
v^0(X(0, w)) = w.
\]
Next, let us examine the characteristic curve \(t \mapsto X(t, w)\) being the solution of
\[
X_t(t, w) = f(v(t, X(t, w))), \quad t \geq 0
\]
\[
X_t(t = 0, w) = X(0, w).
\]
Since \(v\) is constant along its characteristics, we get
\[
v(t, X(t, w)) = w,
\]
from which we extract that
\[
X_t(t, w) = f(v(t, X(t, w))) = f(w),
\]
and therefore,
\[
X(t, w) = X(0, w) + tf(w).
\]
Now we are in a position to calculate
\[
d_p(v, \tilde{v}) = W_p(v_x, \tilde{v}_x)
\]
\[
= \left( \int_0^1 |X(t, w) - \tilde{X}(t, w)|^p \, dw \right)^{\frac{1}{p}}
\]
\[
= \left( \int_0^1 |X(0, w) - \tilde{X}(0, w)|^p \, dw \right)^{\frac{1}{p}}
\]
\[
= W_p(v^0(x), \tilde{v}^0(x))
\]
\[
= d_p(v^0, \tilde{v}^0).
\]
For the general case in which \(v, \tilde{v} \in U\), we define an operator on \(U\) by
\[
T_h v(x) = \int_0^1 \chi \{ X(h, w) \geq x \} (w) \, dw.
\]
Let us recall two important properties of this operator:
**Proposition 1.4.**

(i) \( v \in \mathcal{U}_p \Rightarrow T_h v \in \mathcal{U}_p \)

(ii) For all \( v, \tilde{v} \in \mathcal{U} \),

\[
W_p([T_h v]_x, [T_h \tilde{v}]_x) \leq W_p(v_x, \tilde{v}_x).
\]

**Proof.** See [3]. □

Now one can use this operator to construct an approximate solution \( S_h u^0 \) of the conservation law.

**Definition 1.5.** Let \( h \) be a positive number and \( v \in \mathcal{U} \). For all \( t \geq 0 \) decomposed as \( t = (N + s)h \), \( N \in \mathbb{N} \), and \( s \in [0,1) \), we define

\[
S_h v(t, \cdot) = (1-s)T^N_h v(\cdot) + sT^{N+1}_h v(\cdot),
\]

where \( T^0_h v = v \) and \( T^{N+1}_h = T_h(T^N_h v) \).

Let us also mention the

**Proposition 1.6.** Let \( u^0 \in \mathcal{U}_p \) and \( u \) be the entropy solution to the inviscid scalar conservation law with initial datum \( u^0 \). Then, for any \( t \geq 0 \),

\[
W_p([S_h u^0]_x(t, \cdot), u_x(t, \cdot)) \xrightarrow{h \to 0} 0.
\]

**Proof.** See [3]. □

Finally, we state the contraction property of entropy solutions of the inviscid scalar conservation law:

**Theorem 1.7.** Given a locally Lipschitz, real-valued function \( f \) on \( \mathbb{R} \) and two initial data \( u^0, \tilde{u}^0 \in \mathcal{U}_p \), let \( u, \tilde{u} \) be the corresponding entropy solutions. Then, for any \( t \geq 0 \) and any \( p \geq 1 \), we have, with possibly infinite values,

\[
W_p(u_x(t, \cdot), \tilde{u}_x(t, \cdot)) \leq W_p(u^0_x, \tilde{u}^0_x).
\]

**Proof.** For the ensuing proof, we shall make use of the convergence of the approximate solutions \( S_h u^0(t, \cdot), S_h \tilde{u}^0(t, \cdot) \) to the entropy solutions \( u, \tilde{u} \) with initial conditions \( u^0, \tilde{u}^0 \), respectively, in the \( W_p \)-scheme. We calculate

\[
W_p(u_x(t, \cdot), \tilde{u}_x(t, \cdot)) \leq W_p(u^0_x, \tilde{u}^0_x) + W_p([S_h u^0]_x(t, \cdot), [S_h u^0]_x(t, \cdot)) + W_p([S_h \tilde{u}^0]_x(t, \cdot), \tilde{u}_x(t, \cdot)).
\]

Now the first and the third summand on the right hand side vanish as \( h \to 0 \) as a consequence of Proposition 1.6, and

\[
W_p([S_h u^0]_x(t, \cdot), [S_h \tilde{u}^0]_x(t, \cdot)) \leq W_p(u^0_x, \tilde{u}^0_x)
\]

because of the convexity of the \( W_p \) distance (see Chapter 1, Proposition 3.1). So the proof is complete. □
2. Viscous Scalar Conservation Laws

Here, we would like to fill some gaps the authors in this paper left for proving the analogous result for the viscous case \(9\).

We shall consider solutions in the sense of distributions: A function \(u \in L^\infty([0, +\infty[\times \mathbb{R})\) is a solution of (9) if for all \(\varphi \in C\left([0, +\infty[\times \mathbb{R}\right)\),

\[
\int_0^{+\infty} \int_{-\infty}^{+\infty} (u \varphi_t + f(u) \varphi_x + u \varphi_{xx}) \, dt \, dx + \int_{-\infty}^{+\infty} u^0(x) \varphi(0, x) \, dx = 0.
\]

Existence and uniqueness of this IVP have been established for instance in \([12]\). We also know the subsequent properties:

**Proposition 2.1.**

(i) \(\|u\|_{L^\infty([0, +\infty[\times \mathbb{R})} = \|u^0\|_{L^\infty([0, +\infty[\times \mathbb{R})}\).

(ii) Let \(u^0, v^0\) be in \(L^\infty(\mathbb{R})\) with respective solutions \(u\) and \(v\) to (1). Then for all \(t \geq 0\), we have that

\[
\|u(t, .) - v(t, .)\|_{L^1(\mathbb{R})} \leq \|u^0(x) - v^0(x)\|_{L^1(\mathbb{R})}.
\]

(iii) Let \(u^0, v^0 \in L^\infty\) with \(u^0 \leq v^0\) almost everywhere. Then the respective solutions \(u\) and \(v\) to (1) satisfy \(u(t, x) \leq v(t, x)\) for all \(t \geq 0\) and almost every \(x \in \mathbb{R}\).

(iv) As a consequence, if \(u^0\) is non-decreasing a.e., then so is \(u(t, .)\) for all \(t \geq 0\).

(v) If \(u^0\) has bounded variation, then so has \(u(t, .)\), and for the total variation defined as

\[
TV(u(t, x)) = \lim_{h \to 0} h^{-1} \int_{-\infty}^{+\infty} [u(t, x + h) - u(t, x)] \, dx,
\]

\[
TV(u(t, .)) \leq TV\left(u^0(\cdot)\right).
\]

Notice that properties (i) and (v) are immediate results of (ii), and that (ii) in turn leads directly to (iii), and consequently, (iv), by

\[
0 \leq \|v(t, .) - u(t, .)\|_{L^1(\mathbb{R})} \leq \|u^0(\cdot) - v^0(\cdot)\|_{L^1(\mathbb{R})} = \int_{\mathbb{R}} (u^0(x) - v^0(x)) \, dx = \int_{\mathbb{R}} (v(t, x) - u(t, x)) \, dx,
\]

which is to say that \(v - u \geq 0\) a.e.

**Proof.** For a full proof, see \([12]\). \(\square\)

The following proposition marks the beginning of the new approach by means of Wasserstein distances.

**Proposition 2.2.** Given the initial datum \(u^0 \in \mathcal{U}\), \(u(t, .)\) remains in \(\mathcal{U}\) for all \(t \geq 0\).
2. VISCOS SCALAR CONSERVATION LAWS

Proof. For \( t \geq 0 \), the solution \( u(t, \cdot) \) takes values between 0 and 1, has total variation bounded by 1, is continuous but at countably many points, and almost everywhere non-decreasing. Then its limits at \(-\infty\) and \(+\infty\) are 0 and 1, respectively, since the quantity \( \int_{-\infty}^{+\infty} \mathrm{d}u_x(t, x) \), which amounts to the total variation of \( u(t, \cdot) \), is conserved and equal to 1. \( \Box \)

The equation \( \int_{\mathbb{R}} \mathrm{d}v(x) = TV(v) \) is not hard to get convinced of, yet nevertheless worth noticing.

In order to prove a similar contraction result as for (8), we shall employ a new operator on \( \mathcal{U} \):

\[
(12) \quad T_h u^0 \equiv K_h \ast T_h u^0,
\]

where \( K_h \) stands for the heat kernel on \( \mathbb{R} \) defined by

\[
K_h(z) \equiv \frac{1}{\sqrt{4\pi h}} e^{-z^2/4h}.
\]

Denoting by \( g \) the standard Gaussian on \( \mathbb{R} \) defined as

\[
g(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2},
\]

we can calculate

\[
T_h v(x) = \int_{\mathbb{R}} T_h v(x - z) K_h(z) \, dz = \int_{\mathbb{R}} T_h v(x - \sqrt{2h}y) g(y) \, dy
\]

\[
= \int_{0}^{1} \int T_h v(x - \sqrt{2h}y - hf'(w), w) g(y) \, dw \, dy,
\]

where \( jv : \mathbb{R} \times [0, 1] \to \{0, 1\} \),

\[
 jv(x, w) \equiv \chi \{ v > w \} (x),
\]

thus obtaining the time discretization for (9) introduced in [4].

The new operator \( T_h \) inherits the properties of the former \( T_h \). Indeed, if \( u^0 \in \mathcal{U} \), then \( T_h u^0 \) lies in \( \mathcal{U} \) as well, because it is the convolution of an element in \( \mathcal{U} \) (which is \( T_h u \)) with a Gaussian with unit mass. We see that for \( u \in \mathcal{U}_p \), \( p \geq 1 \), it again belongs to \( \mathcal{U}_p \) by bounding the \( p \)-th moment of its derivative. One can also show that

\[
W_p([T_h u^0]_x, [T_h v^0]_x) \leq W_p(u^0_x, v^0_x).
\]

Proof. Here is a neat proof of this fact F. Bolley pointed out to the author:

Firstly, we give a weaker statement of Chapter 1, Proposition 3.1: For two probability measures \( \mu, \nu \) on the real line,

\[
W_p \left( \frac{\mu_1 + \mu_2}{2}, \frac{\nu_1 + \nu_2}{2} \right) \leq \frac{1}{2} W_p(\mu_1, \nu_1) + \frac{1}{2} W_p(\mu_2, \nu_2).
\]

One can moreover yield the following result by induction:

\[
W_p \left( \sum_{k=1}^{N} \alpha_k \mu_k, \sum_{k=1}^{N} \alpha_k \nu_k \right) \leq \sum_{k=1}^{N} \alpha_k W_p(\mu_k, \nu_k),
\]

where \( \sum_{k=1}^{N} \alpha_k = 1 \), \( \alpha_k \geq 0 \) for all \( k \) and \( (\mu_k)_{k=1}^{N}, (\nu_k)_{k=1}^{N} \) are sequences of probabilities measures on the real line.

At last, consider the aforementioned convolution with the heat kernel \( K_h \).
Since it has unit mass, we can regard it as a limit of the above sequence \((\alpha_k)^N_{k=1}\) as \(N \to \infty\) and generalize the latter case to get:

\[
W_p(K_h \ast \mu(\cdot), K_h \ast \nu(\cdot)) = W_p \left( \int_{-\infty}^{+\infty} K_h(y) \mu(\cdot - y) dy, \int_{-\infty}^{+\infty} K_h(y) \nu(\cdot - y) dy \right)
\]

(14)

\[
\leq \int_{-\infty}^{+\infty} W_p(\mu(\cdot - y), \nu(\cdot - y)) K_h(y) dy
\]

(15)

\[
= \int_{-\infty}^{+\infty} W_p(\mu(\cdot), \nu(\cdot)) K_h(y) dy
\]

Inequality (14) is the generalization of (13), and equality (15) arises from the invariance of Wasserstein distances under translations (think of our pile of sand and the hole it shall be transferred into "shifted" at equal distances). 

Just as in Definition 1.5, with the help of \(T_h\), we build an approximate solution:

**Definition 2.3.** Let \(h \geq 0\) and \(v \in \mathcal{U}\). For all \(t \geq 0\) decomposed as \(t = (N + s)h, N \in \mathbb{N}, s \in [0, 1)\), we define

\[
S_h v(t, \cdot) = (1 - s)T_h^N v(\cdot) + sT_h^{N+1} v(\cdot),
\]

where \(T_h^0 v \equiv v\) and \(T_h^{N+1} v \equiv T_h(T_h^N v)\).

For the Propositions to follow, we shall need the

**Lemma 2.4.** Let \(v \in \mathcal{U}\), twice differentiable with \(v'' \in L^\infty(\mathbb{R})\). Then

\[
\|v \ast K_h - v\|_{L^1(\mathbb{R})} \leq h \|v''\|_{L^1(\mathbb{R})}
\]

and therefore

\[
\|T_h v - v\|_{L^1(\mathbb{R})} \leq h \left[ \|f''\|_{L^\infty((0,1))} + \|v''\|_{L^1(\mathbb{R})} \right].
\]

**Proof.** Take an arbitrary test function \(\varphi \in \mathcal{C}^2_c(\mathbb{R})\). We compute

\[
\int_{-\infty}^{+\infty} \left[ v \ast K_h(x) - v(x) \right] \varphi(x) dx
\]

\[
= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} v(x) K_h(y) [\varphi(x + y) - \varphi(x)] dy dx
\]

\[
= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} v(x) K_h(y) \left[ \int_0^1 \varphi'(x + sy) y dy \right] dx dy
\]

\[
= \int_{0}^{1} \int_{-\infty}^{+\infty} v(x) \left[ \int_{-\infty}^{+\infty} K_h(y) y \varphi'(x + sy) dy \right] dx ds
\]
For the first equality, we have simply performed the change of variables \( x \mapsto x + y \), and for the second, we have used the mean-value theorem. The next step makes use of the fact that \((-2h)K'_h(y) = K_h(y)y\):

\[
(-2h)K'_h(y) = (-2h)\frac{1}{\sqrt{4\pi h}}e^{\frac{-y^2}{2h}} = K_h(y)y.
\]

So we can proceed

\[
1 \int_0^{+\infty} v(x) \left( \int_{-\infty}^{+\infty} K_h(y)\varphi'(x + sy) \, dy \right) \, dx \, ds = \int_0^{+\infty} v(x) \left( \int_{-\infty}^{+\infty} (-2h)K'_h(y)\varphi'(x + sy) \, dy \right) \, dx \, ds
\]

(16)

\[
= \int_0^{+\infty} v(x) \left( \int_{-\infty}^{+\infty} K_h(y)\varphi''(x + sy) \, dy \right) \, dx \, ds
\]

(17)

\[
\leq 2h \int_0^{+\infty} s \int_{-\infty}^{+\infty} K_h(y) dy \|v''\|_{L^1(\mathbb{R})} \|\varphi\|_{L^\infty(\mathbb{R})}
\]

\[
= h \|v''\|_{L^1(\mathbb{R})} \|\varphi\|_{L^\infty(\mathbb{R})},
\]

where for (16), we integrated partially once, and for (17), twice. This establishes the first statement.

For the second one, let us first of all prove that

\[
\|T_h u - u\|_{L^1(\mathbb{R})} \leq \|f'\|_{L^\infty((0,1))} TV(u).
\]

We compute

\[
\|T_h u - u\|_{L^1(\mathbb{R})} \leq \int_0^{+\infty} \int_{-\infty}^{+\infty} |j u(x - h f'(w), w) - j u(x, w)| \, dx \, dw
\]

\[
\leq h \|f'\|_{L^\infty((0,1))} \int_{-\infty}^{+\infty} TV(j u(\cdot, w)) \, dw
\]

by the definition of the total variation \( TV \). Now from the accomplished computations, the identity

\[
TV(u) = \int_0^1 TV(j u(\cdot, w)) \, dw,
\]
and the fact that for all \( u \in \mathcal{U} \), the quantity \( TV(u) \) is at most 1, we finally deduce
\[
\|T_h u - u\|_{L^1(\mathbb{R})} \leq \|T_h u - T_h u\|_{L^1(\mathbb{R})} + \|T_h u - u\|_{L^1(\mathbb{R})}
\]
\[
\leq h \left[ \|u''\|_{L^1(\mathbb{R})} + \|f'\|_{L^\infty([0,1])} \right],
\]
which is the second statement. \( \square \)

**Lemma 2.5.** Let \( u, v \in \mathcal{U} \) such that \( (v - u) \in L^1(\mathbb{R}) \). Then
\[
\|T_h v - T_h u\|_{L^1(\mathbb{R})} \leq \|v - u\|_{L^1(\mathbb{R})},
\]
and hence
\[
\|T_h^N v - T_h^N u\|_{L^1(\mathbb{R})} \leq \|v - u\|_{L^1(\mathbb{R})}
\]
for any \( N \in \mathbb{N} \).

**Proof.** We use Youngs inequality to gain
\[
\|T_h v - T_h u\|_{L^1(\mathbb{R})} = \|K_h * T_h v - K_h * T_h u\|_{L^1(\mathbb{R})}
\]
\[
\leq \|K_h\|_{L^1(\mathbb{R})} \|T_h v - T_h u\|_{L^1(\mathbb{R})} = \|T_h v - T_h u\|_{L^1(\mathbb{R})}
\]
and advance as follows:
\[
\|T_h v - T_h u\|_{L^1(\mathbb{R})}
\]
\[
= \int_{-\infty}^{+\infty} \frac{1}{1} \left| jv(x - hf'(w), w) - ju(x - hf'(w), w) \right| dw \right| dx
\]
\[
\leq \int_{0}^{+\infty} \left( \int_{-\infty}^{+\infty} \left| jv(x - hf'(w), w) - ju(x - hf'(w), w) \right| dx \right) dw
\]
\[
= \int_{0}^{+\infty} \left( \int_{-\infty}^{+\infty} \left| jv(y, w) - ju(y, w) \right| dy \right) dw
\]
\[
= \|v - u\|_{L^1(\mathbb{R})}.
\]
The second statement follows inductively. \( \square \)

Next, we shall prove two contraction properties of these approximate solutions.

**Proposition 2.6.** Let \( h \) be some fixed positive number. Then, for any \( v \in \mathcal{U} \) with \( v'' \in L^1(\mathbb{R}) \) and \( s, t \geq 0 \), we have
\[
\|S_h v(t, \cdot) - S_h v(s, \cdot)\|_{L^1(\mathbb{R})} \leq |t - s| \left[ \|f'\|_{L^\infty([0,1])} + \|v''\|_{L^1(\mathbb{R})} \right].
\]

**Proof.** Lemma 2.4 asserts that
\[
\|T_h V - V\|_{L^1(\mathbb{R})} \leq h \left[ \|f'\|_{L^\infty([0,1])} + \|v''\|_{L^1(\mathbb{R})} \right]
\]
for any \( V \in \mathcal{U} \). Let \( t = (M + \mu)h \) and \( s = (N + (1 - \nu))h \) with \( M, N \in \mathbb{N}, 0 \leq \mu < 1, 0 < \nu \leq 1 \), and, without loss of generality, assume that \( M > N \). We prove the
special case $\mu = 0, \nu = 1$ first.

So

$$
\|S_h v(Mh, .) - S_h v(Nh, .)\|_{L^1(\mathbb{R})}
= \|T_h^M v(.) - T_h^N v(.)\|_{L^1(\mathbb{R})}
\leq \|T_h^M v(.) - T_h^{M-1} v(.)\|_{L^1(\mathbb{R})} + \ldots + \|T_h^{N+1} v(.) - T_h^N v(.)\|_{L^1(\mathbb{R})}
= \|T_h^{M-1} (T_h v(.) - T_h^{M-1} v(.)\|_{L^1(\mathbb{R})} + \ldots + \|T_h^N (T_h v(.) - T_h^N v(.)\|_{L^1(\mathbb{R})}
\leq (M - N) h \left[ \|f'\|_{L^\infty([0,1])} + \|v''\|_{L^1(\mathbb{R})} \right]
= |t - s| \left[ \|f'\|_{L^\infty((0,1))} + \|v''\|_{L^1(\mathbb{R})} \right],
$$

where we have used Lemmata 2.4, 2.5 for the second inequality.

Next, let us proceed to the general case, i.e. choose $\mu, \nu \in (0, 1)$. Then

$$
\|S_h v(t, .) - S_h v(s, .)\|_{L^1(\mathbb{R})}
= \|T_h^M v(.) + \mu T_h^{M+1} v(.) - (\nu T_h^N v(.) + (1 - \nu) T_h^{N+1} v(.)\|_{L^1(\mathbb{R})}
\leq \|T_h^M v(.) - T_h^{N+1} v(.)\| + \mu \|T_h^{M+1} v(.) - T_h^M v(.)\|
+ \mu \|T_h^{N+1} v(.) - T_h^N v(.)\|
\leq (M - N - 1 + \mu + \nu) h \left[ \|f'\|_{L^\infty([0,1])} + \|v''\|_{L^1(\mathbb{R})} \right]
= |t - s| \left[ \|f'\|_{L^\infty((0,1))} + \|v''\|_{L^1(\mathbb{R})} \right].
$$

Hence the proof is complete. \( \square \)

From this bound we infer, as in the inviscid case, the following

**Proposition 2.7.** Let $u \in \mathcal{U}$ be twice differentiable with $u'' \in L^1(\mathbb{R})$. Then the sequence $(S_h u)_h$ is relatively compact in $C \left( [0, +\infty[, L^1_{\text{loc}}(\mathbb{R}) \right)$. 

In order to prove that the limit of any converging subsequence of $(S_h u^0)_h$ in $C \left( [0, +\infty[, L^1_{\text{loc}}(\mathbb{R}) \right)$ is a solution of (9), we first prove the

**Proposition 2.8.** Let $u \in \mathcal{U}$ and $\varphi \in C_c^\infty(\mathbb{R})$ with its compact support being contained in $[-R, R]$. Then, for $M = \|f'\|_{L^\infty((0,1))}$, we have

$$
\left| \int_{-\infty}^{+\infty} \left[ T_h v(x) - v(x) \right] \varphi(x) dx - \int_{-\infty}^{+\infty} v(x) \varphi''(x) dx - h \int_{-\infty}^{+\infty} f(v(x)) \varphi'(x) dx \right|
\leq h^2 \left[ M^2 (R^2 - hM) + M \right] \|\varphi''\|_{L^\infty([-R,R])} + \frac{4}{3\sqrt{\pi}} h \sqrt{h \|\varphi''\|_{L^\infty([-R,R])}}.
$$
PROOF. The result can be concluded from the following three estimates:

\[ \|T_h v - v\|_{L^1(\mathbb{R})} \leq hM, \]

\[ \left| \int_{-\infty}^{+\infty} [T_h v(x) - v(x)] \varphi(x) \, dx - h \int_{-\infty}^{+\infty} f(v(x)) \varphi'(x) \, dx \right| \leq h^2 M^2 (R + hM) \|\varphi''\|_{L^\infty([-R,R])}, \]

and

\[ \left| \int_{-\infty}^{+\infty} [K_h * (v(x) - v(x))] \varphi(x) \, dx - h \int_{-\infty}^{+\infty} v(x) \varphi'' \, dx \right| \leq \frac{4}{3\sqrt{\pi}} h^{\frac{1}{2}} \|\varphi''\|_{L^\infty([-R,R])}. \]

We have already encountered the first inequality; the second can be derived from by arguments similar to those used in the proof of Lemma 2.4, and the third one follows from expanding

\[ \varphi(x + y) = \varphi(x) + h\varphi'(x) + \frac{y^2}{2}\varphi''(x) + \frac{y^3}{3!}\int_0^1 \varphi^{(3)}(x + ty)(1 - t)^2 \, dt \]

and noting that \( \int_{-\infty}^{+\infty} yK_h(y) \, dy = 0 \) since \( K_h \) is odd. This last estimate is then applied to \( T_h v \) instead of \( v \) in order to obtain the result of this Proposition. \( \square \)

Moreover, we have the ensuing \( W_p \)-contraction property:

PROPOSITION 2.9. For fixed \( h \geq 0 \), and \( v, \tilde{v} \in \mathcal{U} \), we have for all \( t \geq 0 \) that

\[ W_p([S_h v]_x, [S_h \tilde{v}]_x) \leq W_p(v_x, \tilde{v}_x). \]

PROOF. The statement is due to the convexity of the Wasserstein distance of order \( p \), see Proposition 2.4. \( \square \)

PROPOSITION 2.10. Let \( u^0 \in \mathcal{U} \) with \( (u^0)''' \in L^1(\mathbb{R}) \). Then, as \( h \) tends to 0, the sequence \( (S_h u^0)_h \) converges to the solution of the viscous scalar conservation law with initial datum \( u^0 \) in \( C([0, +\infty[, L^1_{\text{loc}}(\mathbb{R})] \).

PROOF. It is a fact that the viscous scalar conservation law has a unique solution in the sense of distributions. We also know that the family \( (S_h u^0)_h \) is relatively compact in \( C([0, +\infty[, L^1_{\text{loc}}(\mathbb{R})] \) due to Proposition 2.4 and the theorems of Helly and Arzelà-Ascoli. By Proposition 2.9, it can be checked that the limit of any subsequence of \( (S_h u^0)_h \) tends to the solution of the viscous scalar conservation law with twice differentiable initial datum in \( C([0, +\infty[, L^1_{\text{loc}}(\mathbb{R})] \). By the uniqueness of this solution, this concludes the proof. \( \square \)

THEOREM 2.11. Given a locally Lipschitz, real-valued function \( f \) on \( \mathbb{R} \) and two initial data \( u^0, \tilde{u}^0 \) in \( \mathcal{U} \), and let \( u, \tilde{u} \) be the corresponding solutions to (19). Then, for any \( t \geq 0 \) and any \( p \geq 1 \), we have (with possibly infinite values):

\[ W_p(u_x(t, .), \tilde{u}_x(t, .)) \leq W_p(u^0_x(\cdot), \tilde{u}^0_x(\cdot)). \]
PROOF. We do this proof in two steps.

1. Given \( t \geq 0 \), and assuming that the initial data \( u^0, \tilde{u}^0 \) lie in \( U \) and are twice differentiable with second derivatives in \( L^1 \), Propositions 2.8 and 2.10 yield the convergence in \( L_{1, \infty}^1(\mathbb{R}) \) of \( (S_h u^0(t, \cdot))_h \) to the solution \( u(t, \cdot) \), so \( ([S_h u^0]_x(t, \cdot))_h \) tends to \( u_x(t, \cdot) \) in the weak sense of probability measures, that is, for all \( \varphi \in \mathcal{C}_b(\mathbb{R}) \),

\[
\lim_{h \to 0^+} \int_{-\infty}^{+\infty} \varphi(x) d\left([S_h u^0]_x(t, x)\right) = \int_{-\infty}^{+\infty} \varphi(x) du_x(t, x).
\]

Since this also holds for \( \tilde{u}^0 \), we are in a position to compute

\[
W_p(u_x(t, \cdot), \tilde{u}_x(t, \cdot)) \leq \liminf_{h \to 0^+} W_p \left([S_h u^0]_x, [S_h \tilde{u}^0]_x\right),
\]

as a consequence of weak convergence.

Now, for each \( h \),

\[
W_p \left([S_h u^0]_x(t, \cdot), [S_h \tilde{u}^0]_x(t, \cdot)\right) \leq W_p \left(u_x^0, \tilde{u}_x^0\right),
\]

so in the end, we get

\[
W_p(u_x(t, \cdot), \tilde{u}_x(t, \cdot)) \leq W_p \left(u_x^0, \tilde{u}_x^0\right).
\]

This concludes the argument for this case.

2. Here we extend the theorem to the case of arbitrary initial data \( u^0, \tilde{u}^0 \) in \( U \). To this end, we shall take mollifiers \( (\rho^\varepsilon)_{\varepsilon > 0} \) defined on \( \mathbb{R} \) by

\[
\rho^\varepsilon(x) \equiv \frac{1}{\varepsilon} \rho \left(\frac{x}{\varepsilon}\right),
\]

where \( \rho \) represents a non-negative \( C^\infty(\mathbb{R}) \)-function with unit mass.

For any \( \varepsilon > 0 \) and \( v \in U \), we define \( v^\varepsilon \equiv \rho^\varepsilon \ast v \). Then \( v^\varepsilon \) is twice differentiable with \( (v^\varepsilon)' \in L^1(\mathbb{R}) \), plus it belongs to \( U_p, p \geq 1 \), if \( v \) does, because

\[
\int_{-\infty}^{+\infty} |x|^p \, dv^\varepsilon_x(x) = \int_{-\infty}^{+\infty} |x|^p \, d \left[\rho^\varepsilon \ast v\right]_x(x) = \int_{-\infty}^{+\infty} |x|^p \, d \left[ \int_{-\infty}^{+\infty} \rho^\varepsilon(y) v(x - y) \, dy \right]_x(x) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |x|^p \rho^\varepsilon(y) \, dv_x(x - y) \, dy = \int_{-\infty}^{+\infty} \rho^\varepsilon(y) \int_{-\infty}^{+\infty} |x|^p \, dv_x(x - y) \, dy < +\infty.
\]

The finiteness of the last term follows from the the boundedness of the \( p^{th} \) moments of \( v \in U_p \), and the unit mass of the mollifier \( \rho^\varepsilon \). Moreover, we know that

\[
\|v^\varepsilon - v\|_{L^1(\mathbb{R})} \leq \varepsilon \int_{-\infty}^{+\infty} |z| \rho(z) \, dz,
\]
as

\[ TV(v) \equiv \lim_{y \to 0} \left| y \right|^{-1} \int_{-\infty}^{+\infty} \left[ v(x + y) - v(y) \right] \ dx = 1 \]

\[ \Rightarrow \int_{-\infty}^{+\infty} [v(x + y) - v(y)] \ dx \leq |y|, \]

for all \( v \in \mathcal{U} \), and therefore

\[ \|v^\varepsilon - v\|_{L^1(\mathbb{R})} = \int_{-\infty}^{+\infty} v^\varepsilon(x) - v(x) \ dx \]

\[ = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \left( \rho^\varepsilon(x - y) v(y) \right) \ dy - v(x) \ dx \]

\[ = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \rho^\varepsilon(y) \left[ v(x + y) - v(y) \right] \ dy \ dx \leq \int_{-\infty}^{+\infty} \rho^\varepsilon(y) |y| \ dy \]

\[ = \int_{-\infty}^{+\infty} \rho \left( \frac{y}{\varepsilon} \right) |y| \ dy = \int_{-\infty}^{+\infty} \frac{1}{\varepsilon} \rho(z) |z| \ dz = \varepsilon \int_{-\infty}^{+\infty} |z| \rho(z) \ dz. \]

Now for two initial data \( u^0, \tilde{u}^0 \) in \( \mathcal{U} \) with respective solutions \( u, \tilde{u} \), the hypotheses of the first step are valid for the mollified \( u^0_{\varepsilon}, \tilde{u}^0_{\varepsilon} \), so by denoting by \( u^\varepsilon \) and \( \tilde{u}^\varepsilon \) their respective solutions, we already know that

\[ W_p (u^\varepsilon_x(t, \cdot), \tilde{u}^\varepsilon_x(t, \cdot)) \leq W_p (u^0_{\varepsilon}(\cdot), \tilde{u}^0_{\varepsilon}(\cdot)). \]

Furthermore,

\[ u^0_x = u^0_{\varepsilon} * \rho^\varepsilon \]

\[ \tilde{u}^0_x = \tilde{u}^0_{\varepsilon} * \rho^\varepsilon \]

\[ \int_{-\infty}^{+\infty} \rho^\varepsilon(x) \ dx = 1, \]

so by the convexity of \( W_p \), we get

\[ W_p (u^0_{\varepsilon}(\cdot), \tilde{u}^0_{\varepsilon}(\cdot)) \leq W_p (u^0_{\varepsilon}(\cdot), \tilde{u}^0_{\varepsilon}(\cdot)). \]

The classical \( L^1 \)-contraction property yields

\[ \|u^\varepsilon(t, \cdot) - u(t, \cdot)\|_{L^1(\mathbb{R})} \leq \|u^0_{\varepsilon} - u^0\|_{L^1(\mathbb{R})}, \]

and consequently, we see that

\[ \|u^0_{\varepsilon} - u^0\|_{L^1(\mathbb{R})} \leq \varepsilon \int_{-\infty}^{+\infty} |z| \rho(z) \ dz. \]

Particularly, \( u^\varepsilon(t, \cdot) \) converges to \( u(t, \cdot) \) in \( L^1_{loc}(\mathbb{R}) \). Hence \( u^\varepsilon_x(t, \cdot) \) tends to \( u_x(t, \cdot) \) in the weak sense of probability measures, which as we already know implies

\[ W_p (u_x(t, \cdot), \tilde{u}_x(t, \cdot)) \leq \lim_{\varepsilon \to 0} \inf W_p (u^\varepsilon_x(t, \cdot), \tilde{u}^\varepsilon_x(t, \cdot)). \]
so finally, we get

\[ W_p \left(u_x(t, \cdot), \tilde{u}_x(t, \cdot) \right) \leq W_p \left(u^0_x, \tilde{u}^0_x \right) \]

by the estimates from above. Hence the proof is complete. \qed
CHAPTER 3

$W_{2n}$-Stability for Drift-Diffusion-Poisson Systems.

In this chapter, we will further investigate applications of Wasserstein distances to nonlinear PDEs, namely the large time behavior of drift-diffusion-Poisson (ddP) systems that appear both in semiconductor device modeling and in plasma physics. As a reference, consult e. g. [2].

1. ddP-Systems with Linear Diffusion

The first case is a toy model in one space dimension for the carrier transport in semiconductor devices:

\begin{align*}
\rho_t &= \frac{\partial}{\partial x} \left( \rho_x + \rho (V(x) + \psi)_x \right) \\
-\psi_{xx} &= \rho - C(x) \\
\rho(x, t = 0) &= \rho_0(x) \geq 0.
\end{align*}

$\rho(x, t)$ denotes the spatial distribution of, for instance, the negatively charged electrons, and $\psi(x, t)$ stands for the self-consistent electrostatic potential created by the charge carrier and the doping profile $C(x)$. Concerning the initial datum we require that $\rho_0(x)$ lie in $L^1(\mathbb{R}) \cap L^\infty(\mathbb{R})$ and that $\|\rho_0\|_{L^1(\mathbb{R})} = 1$, so we can actually interpret $\rho_0$ as a probability density function. As a result of the divergence form of equation (18), the solutions remain probability density functions for all times.

Let us now perform a standard procedure. First of all, introduce the distribution function of $\rho$ (in other words, its primitive):

$$F(x, t) \equiv \int_{-\infty}^x \rho(y, t) dy.$$  

Then,

$$F_t(x, t) = \int_{-\infty}^x \frac{\partial}{\partial y} \left( \rho_y + \rho (V(y) + \psi)_y \right) dy = F_{xx}(x, t) + F_x(x, t)V'(x) + F_x(x, t)\psi_x(x, t).$$

Let us now examine the equation for the pseudo-inverse $F^{-1}$ associated to the original system. Since

$$F(F^{-1}(\xi, t), t) = \xi,$$

for all $\xi \in [0, 1]$, we can compute

$$\frac{\partial}{\partial \xi} [F(F^{-1}(\xi, t), t)] = \frac{\partial F}{\partial x}|_{x=F^{-1}(\xi, t)} \frac{\partial F^{-1}}{\partial \xi}(\xi, t) = 1.$$
and consequently have
\[
\frac{\partial^2}{\partial \xi^2} \left[ F(F^{-1}(\xi, t), t) \right] = \frac{\partial^2 F}{\partial x^2}\big|_{x=F^{-1}(\xi, t)} + \left[ \frac{\partial F^{-1}(\xi, t)}{\partial \xi} \right]^2 F\left( F_1(x; t) \right) = \frac{\partial F}{\partial x}\big|_{x=F^{-1}(\xi, t)} \frac{\partial^2 F^{-1}(\xi, t)}{\partial \xi^2} (\xi, t) = 0.
\]
Let us also calculate the time derivative of (20). It reads
\[
\frac{\partial}{\partial t} \left[ F(F^{-1}(\xi, t), t) \right] = \frac{\partial F}{\partial x}\big|_{x=F^{-1}(\xi, t)} \frac{\partial F^{-1}(\xi, t)}{\partial t} (\xi, t) + \frac{\partial F}{\partial t}(F^{-1}(\xi, t), t) = 0.
\]
Thus we obtain the following (shorthand) equation satisfied by the pseudo-inverse \( F_1 \):
\[
F_1 = \frac{F_1}{F_x} \big|_{x=F^{-1}(\xi, t)} = -\frac{F_{xx}}{F_x} \big|_{x=F^{-1}(\xi, t)} - V'(x) - \psi_x(x, t)
\]
At this point, we dedicate ourselves to the role of the electrostatic potential \( \psi \). Green’s function for the real line gives
\[
-\psi_x\big|_{x=F^{-1}(\xi, t)} = -\frac{\partial}{\partial x}\big|_{x=F^{-1}(\xi, t)} \int_{-\infty}^{+\infty} |x-y| \left[ \rho(y, t) - C(y) \right] dy
\]
\[
= \int_{-\infty}^{+\infty} \text{sign}(x-y) \left[ \rho(y, t) - C(y) \right] dy
\]
\[
= \int_{0}^{1} \text{sign} \left[ F^{-1}(\xi, t) - F^{-1}(\eta, t) \right] \int_{0}^{1} \text{sign} \left[ F^{-1}(\xi, t) - F^{-1}(\eta, t) \right] C \left( F^{-1}(\eta) \right) F^{-1}(\eta) d\eta
\]
\[
(21) = -\int_{0}^{1} \text{sign}(\xi - \eta) d\eta + \int_{0}^{1} \text{sign}(\xi - \eta) C \left( F^{-1}(\eta) \right) F^{-1}(\eta) d\eta.
\]
In line (21), we performed the change of variables for \( \xi, \eta \):
\[
x = F^{-1}(\xi, t) \quad \Rightarrow \quad \xi = F(x, t)
\]
\[
y = F^{-1}(\eta, t) \quad \Rightarrow \quad \eta = F(y, t)
\]
and we get
\[
d\eta = F(y, t) dy = \rho(y, t) dy.
\]
Moreover, line (22) comes from the fact that both pseudo-inverses are increasing.
Let us scrutinize the very last summand of the above computation. Introducing the primitive of \( C \), \( C(\zeta) \equiv \int_{-\infty}^{\zeta} C(z) dz \), we can compute
\[
\frac{\partial}{\partial \eta} \left[ C \left( F^{-1}(\eta, t) \right) \right] = C \left( F^{-1}(\eta, t) \right) F^{-1}(\eta, t),
\]
and as a result, that summand reads
\[
\int_0^1 \text{sign} (\xi - \eta) \frac{\partial}{\partial \eta} C \left( F^{-1}(\eta, t) \right) \, d\eta
\]
\[
= - \frac{\partial}{\partial \eta} \text{sign} (\xi - \eta) C \left( F^{-1}(\eta, t) \right) \, d\eta + \text{sign} (\xi - \eta) C \left( F^{-1}(\eta, t) \right) \bigg|_{\eta=0}^{\eta=1}
\]
\[
= 2 \int_0^1 \delta_\xi(\eta) C \left( F^{-1}(\eta, t) \right) \, d\eta - \int_{-\infty}^{+\infty} C(x) \, dx
\]
\[
= 2C \left( F^{-1}(\xi, t) \right) - \int_{-\infty}^{+\infty} C(x) \, dx.
\]

We collect the terms for the time derivative of the electrostatic potential \(\psi\):
\[
-\psi_x = - \int_0^1 \text{sign} (\xi - \eta) \, d\eta + 2C \left( F^{-1}(\xi, t) \right) - \int_{-\infty}^{+\infty} C(x) \, dx.
\]

So in its final constitution, our new equation looks like
\[
\frac{\partial F^{-1}}{\partial t} (\xi, t) = - \frac{\partial}{\partial \xi} \left[ \frac{1}{F^{-1}_\xi(\xi, t)} \right] - V' \left( F^{-1}(\xi, t) \right)
\]
\[
- \int_0^1 \text{sign} (\xi - \eta) \, d\eta + 2C \left( F^{-1}(\xi, t) \right) - \int_{-\infty}^{+\infty} C(x) \, dx.
\]

(23)

It is very natural to ask now how two different solutions of (23) behave in time in terms of "Wasserstein distances" (cf. Chapter 2, (10)) of order \(2n, n \in \mathbb{N}\). So

(24)
\[
\frac{d}{dt} d_{2n}(F, G)^{2n}
\]
\[
= \frac{d}{dt} \int_0^1 \left| F^{-1}(\xi, t) - G^{-1}(\xi, t) \right|^{2n} \, d\xi
\]
\[
= 2n \int_0^1 \left( F^{-1}(\xi, t) - G^{-1}(\xi, t) \right)^{2n-1} \left[ F^{-1}_\xi(\xi, t) - G^{-1}_\xi(\xi, t) \right] \, d\xi
\]
\[
= 2n \int_0^1 \left( F^{-1}(\xi, t) - G^{-1}(\xi, t) \right)^{2n-1} \left\{ - \frac{\partial}{\partial \xi} \left( \frac{1}{F^{-1}_\xi(\xi, t)} - \frac{1}{G^{-1}_\xi(\xi, t)} \right) 
\right.
\]
\[
\left. - \left( V' \left( F^{-1}(\xi, t) \right) - V' \left( G^{-1}(\xi, t) \right) \right) \right\}
\]
\[
+ 2 \left[ C \left( F^{-1}(\xi, t) \right) - C \left( G^{-1}(\xi, t) \right) \right] \, d\xi,
\]

since the \(- \int_0^1 \text{sign}(\xi - \eta) \, d\eta\) and \(- \int_{-\infty}^{+\infty} C(x) \, dx\)-terms vanish. In order to keep things better available to comprehension, let us investigate the behavior of each of
the summands above. First of all, we calculate:

\[
2n \int_0^1 \left[ F^{-1}(\xi, t) - G^{-1}(\xi, t) \right]^{2n-1} \left[ -\frac{\partial}{\partial \xi} \left( \frac{1}{F^{-1}_\xi(\xi, t)} - \frac{1}{G^{-1}_\xi(\xi, t)} \right) \right] d\xi
\]

\[
= 2n(2n-1) \int_0^1 \left[ F^{-1}(\xi, t) - G^{-1}(\xi, t) \right]^{2n-2} \left[ \frac{1}{F^{-1}_\xi(\xi, t)} - \frac{1}{G^{-1}_\xi(\xi, t)} \right] d\xi
\]

\[
-2n \left[ F^{-1}(\xi, t) - G^{-1}(\xi, t) \right]^{2n-1} \left[ \frac{1}{F^{-1}_\xi(\xi, t)} - \frac{1}{G^{-1}_\xi(\xi, t)} \right]^{\xi=1}_{\xi=0} \leq 0.
\]

We wish to say at this point that this very last estimate can be proven by an approximation argument similar to the one in [5], which enables us to get rid of the boundary term in the last expression. More precisely, one should approximate the general solution by a cutoff sequence of regularized solutions. The detailed proof shall be presented in a forthcoming paper. Nevertheless, we can state that:

\[
\lim_{\xi \to 0^+} \left( F^{-1}_\xi(\xi, t) \right)^{-1} = \lim_{x \to +\infty} \rho(x, t) = 0
\]

\[
\lim_{\xi \to 1^-} \left( F^{-1}_\xi(\xi, t) \right)^{-1} = \lim_{x \to -\infty} \rho(x, t) = 0,
\]

and because of the inequality \((a-b)/(1/a-1/b) = -(a-b)^2/(ab) \leq 0\) for all \(a, b \neq 0\).

Let us now advance to the second summand.

\[
-2n \int_0^1 \left[ F^{-1}(\xi, t) - G^{-1}(\xi, t) \right]^{2n-1} \left[ V'' \left( F^{-1}(\xi, t) \right) - V'' \left( G^{-1}(\xi, t) \right) \right] d\xi
\]

\[
\leq -2n \inf_{x \in \mathbb{R}} V''(x) \int_0^1 \left[ F^{-1}(\xi, t) - G^{-1}(\xi, t) \right]^{2n} d\xi.
\]

Finally, we proceed to the last summand:

\[
4n \int_0^1 \left[ F^{-1}(\xi, t) - G^{-1}(\xi, t) \right]^{2n-1} \left[ C \left( F^{-1}(\xi, t) \right) - C \left( G^{-1}(\xi, t) \right) \right] d\xi
\]

\[
\leq 4n \|C\|_{\text{Lip}} \int_0^1 \left[ F^{-1}(\xi, t) - G^{-1}(\xi, t) \right]^{2n} d\xi
\]

\[
= 4n \|C\|_{\text{Lip}} \int_0^1 \left[ F^{-1}(\xi, t) - G^{-1}(\xi, t) \right]^{2n} d\xi.
\]

So what do all these computations tell us? Recall from [24] that we investigated the time behavior of the "Wasserstein distances" of even orders for distribution functions \(F, G\):

\[
\frac{d}{dt} d_{2n}(F, G)^{2n} \leq \left[ -2n \inf_{x \in \mathbb{R}} V'' + 4n \|C\|_{\text{Lip}} \right] d(F, G)^{2n}.
\]

Also recollect the injection Chapter 2, (14), so that we can understand our results in terms of the "proper" Wasserstein distances of even orders of two solutions \(\rho_1, \rho_2\).
of (18) with respective pseudo-inverses $F, G$:

\[
\frac{d}{dt} W_{2n}(\rho_1, \rho_2)^{2n} \leq \left[ -2n \inf_{x \in \mathbb{R}} V'' + 4n \|C\|_{L^\infty} \right] W_{2n}(\rho_1, \rho_2)^{2n}.
\]

Finally, we take the $2n$-th square root on both sides of (25) and yield the subsequent theorem by Gronwall’s Lemma:

**Theorem 1.1.** Let $\rho_1, \rho_2$ be two solutions of (18), (19) with initial conditions $\rho_1^0$ and $\rho_2^0$ as described, respectively. Then, for all $n \in \mathbb{N}$,

\[
W_{2n}(\rho_1 (., t), \rho_2 (., t)) \leq \exp \left[ t \left( -\inf_{x \in \mathbb{R}} V'' + 2 \|C\|_{L^\infty(\mathbb{R})} \right) \right] W_{2n}(\rho_1^0 (.), \rho_2^0 (.) )
\]

for all $t \geq 0$.

**Remark 1.2.** This implies that whenever

\[
2 \|C\|_{L^\infty(\mathbb{R})} \leq \inf_{x \in \mathbb{R}} V''(x),
\]

we obtain exponential convergence in time.

### 2. ddP-Systems with General Nonlinear Diffusion

It is curious to notice that plugging a non-decreasing diffusion function $\Phi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ with $\Phi(0) = 0$ into the original equation (18) conserves the contraction property in Wasserstein sense.

\[
\rho_t = \frac{\partial}{\partial x} [\Phi(\rho)_x + (V(x) + \psi)_x]
\]

(27)

$$\rho_t = \frac{\partial \psi_{xx}}{\partial x} = \rho - C(x)$$

$$0 \leq \rho(x, 0) = \rho_0(x) \in L^1(\mathbb{R}) \cap L^\infty(\mathbb{R})$$

with $\|\rho_0\|_{L^1(\mathbb{R})} = 1$.

Computing formally the equation satisfied by the pseudo-inverse of the distribution function of the equation

\[
\rho_t = (\Phi(\rho))_{xx}
\]

$$0 \leq \rho(x, 0) = \rho_0(x) \in L^1(\mathbb{R}) \cap L^\infty(\mathbb{R})$$

with $\|\rho_0\|_{L^1(\mathbb{R})}$, we have as usual

\[
F(x, t) = \int_{-\infty}^{x} \rho(y, t) \, dy \Rightarrow \frac{\partial F}{\partial x} |_{x = F^{-1}(\xi, t)} \frac{\partial F^{-1}}{\partial \xi}(\xi, t) = 1
\]

$$\Rightarrow \quad F_x |_{x = F^{-1}(\xi, t)} = \left( F_{\xi}^{-1}(\xi, t) \right)^{-1}$$

$$\Rightarrow \quad F_{\xi}(x, t) = \int_{-\infty}^{x} \frac{\partial^2}{\partial y^2} \Phi(F_y) \, dy = \frac{\partial}{\partial x} [\Phi(F_x(x, t))]$$

$$= \frac{\partial}{\partial x} \left( \Phi \left[ \left( F_{\xi}^{-1}(\xi, t) \right)^{-1} \right] \right).$$

By the chain rule and the definition of $F$, we can write

$$\frac{\partial}{\partial x} |_{x = F^{-1}(\xi, t)} = \frac{\partial \xi}{\partial \xi} \frac{\partial}{\partial \xi} |_{\xi = F(x, t)} = \rho(x, t) \frac{\partial}{\partial \xi},$$
so finally, we get

\[ \frac{\partial F^{-1}}{\partial \xi}(\xi, t) = -\frac{F_t}{F_x}|_{x=F^{-1}(\xi, t)} = -\frac{\partial}{\partial \xi} \left( \Phi \left[ \left( F^{-1}_\xi(\xi, t) \right)^{-1} \right] \right). \]

Again, the Wasserstein distance of even order between two solutions decays in time:

\[ \frac{d}{dt} W_{2n}(\rho_1(x, t), \rho_2(x, t))^{2n} = \frac{d}{dt} \int_0^1 \left[ F^{-1}_\xi(\xi, t) - G^{-1}_\xi(\xi, t) \right]^{2n} d\xi \]

\[ = -2n \int_0^1 \left[ F^{-1}_\xi(\xi, t) - G^{-1}_\xi(\xi, t) \right]^{2n-1} \times \left[ \Phi \left[ \left( F^{-1}_\xi(\xi, t) \right)^{-1} \right] - \Phi \left[ \left( G^{-1}_\xi(\xi, t) \right)^{-1} \right] \right] d\xi \]

\[ = 2n(2n-1) \int_0^1 \left[ F^{-1}_\xi(\xi, t) - G^{-1}_\xi(\xi, t) \right]^{2n-2} \times \left[ \Phi \left[ \left( F^{-1}_\xi(\xi, t) \right)^{-1} \right] - \Phi \left[ \left( G^{-1}_\xi(\xi, t) \right)^{-1} \right] \right] d\xi \]

because \( \Phi \) is non-decreasing, and therefore, the second factor of the integrand in the last integral will always have the opposite sign of the first factor. Moreover, observe that the boundary term vanishes due to a similar approximation procedure as indicated for the linear diffusion and the fact that \( \Phi(0) = 0 \).

Yet for the nonlinear diffusion, we can even let \( n \) tend to infinity in Theorem 1.1 to achieve:

\[ W_\infty(\rho_1(., t), \rho_2(., t)) \leq W_\infty (\rho^0_1(.), \rho^0_2(.)) e^{-\lambda t}, \lambda \geq 0 \]

for the diffusion part in (26).

3. **ddP-Systems for Plasmas**

The drift-diffusion-Poisson system for plasmas modeling the transport of charged ions reads

\[ \begin{align*}
  u_t &= \frac{\partial}{\partial x} (u_x + u \phi_x) \\
  \phi_{xx} &= u \\
  0 \leq u(x, 0) &= u_0(x) \in L^1(\mathbb{R}) \cap L^\infty(\mathbb{R})
\end{align*} \tag{28} \tag{29} \]

with \( \|u_0\|_{L^1(\mathbb{R})} = 1 \).

Following the lines of the previous computations, we get

\[ \frac{\partial F^{-1}}{\partial t}(\xi, t) = -\frac{\partial}{\partial \xi} \left[ \frac{1}{F^{-1}_\xi(\xi, t)} \right] - \int_0^1 \text{sign}(\xi - \eta) \, d\eta, \]

so the \( W_{2n} \)-stability result (for all \( n \in \mathbb{N} \)) follows from Theorem 1.1.
Bibliography