Warwick Mathematics Institute, Ergodic Theory and Dynamical Systems seminar

# On the Wasserstein distance between stationary probability measures

**14 FEB, 2019**

Italo Cipriano

Facultad de Matemáticas Pontificia Universidad Católica de Chile
Proyecto Anillo ACT172001 PIA 583-17

# Motivation

**A list of questions in a paper that Jon Fraser wrote when was here!**

First and second moments for self-similar couplings and Wasserstein distances


Jonathan M. Fraser

*Mathematics Institute, Zeeman Building,*
*University of Warwick, Coventry, CV4 7AL, UK*
*e-mail: jon.fraser32@gmail.com*


January 29, 2014

# Fundamental definitions

- Iterated Function Systems
- Stationary probability measures
- Wasserstein Distances

# Iterated Function System

**Is a finite set of contractions** $f_1, f_2, \ldots, f_N$ **in a complete compact metric space** $\mathcal{X}$ **.**

**Hutchinson proved in [Hut1981] that when** $\mathcal{X} = \mathbb{R}^n$**, there exists a unique non empty compact invariant set, that is,** $\mathcal{S} = \cup_{i=1}^{N} f_i(\mathcal{S})$ **. This set is called <u>attractor</u>.**

**Examples of attractors in** $\mathbb{R}^2$**.**

[Hut1981] J. Hutchinson, Fractals and Self Similarity, Indiana Univ. Math. J. 30 (5): 713-747 (1981), doi:10.1512/iumj.1981.30.30055

# Stationary probability measure

**Given an iterated function systems and a probability vector** $(p_1, p_2, \ldots, p_N)$**,**
**there exists a unique regular Borel probability measure such that**

$$\mu(A) = \sum_{i=1}^{N} p_i \mu(f_i^{-1} A) \text{ for every } A \in \mathscr{B},$$

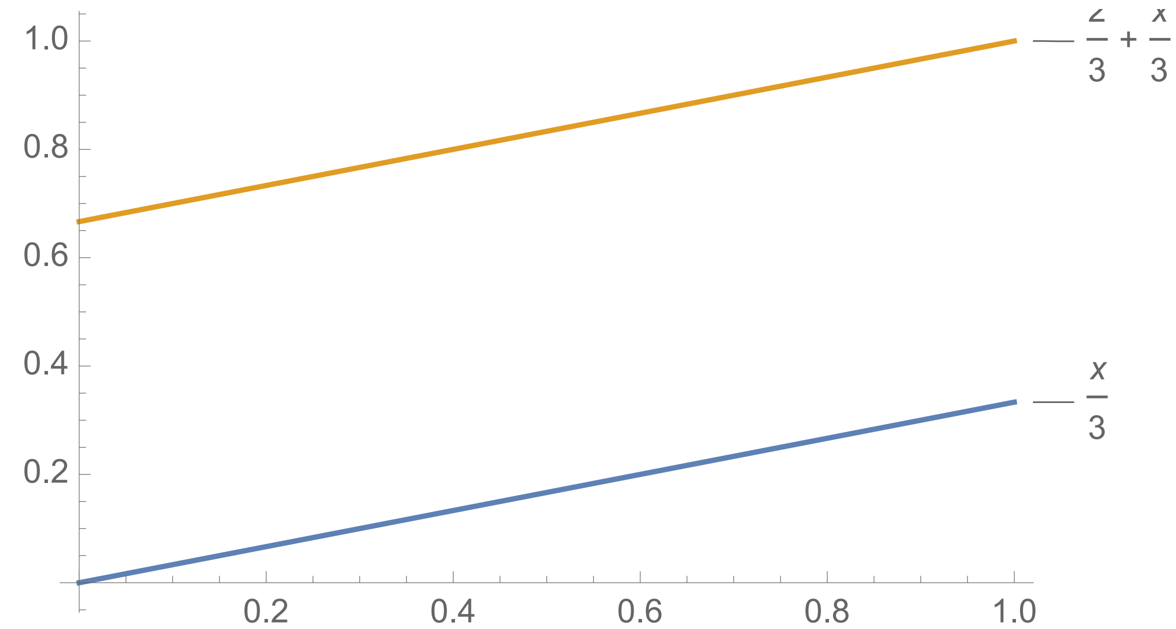
**where** $\mathscr{B}$ **are the subset of Borel of** $\mathbb{R}^N$**.**

**This probability measure is called <u>stationary probability measure</u> and its existence and unicity is proved in [Hut1981].**

# Example

$$f_1(x) = \frac{x}{3}$$

$$f_2(x) = \frac{x}{3} + \frac{2}{3}$$

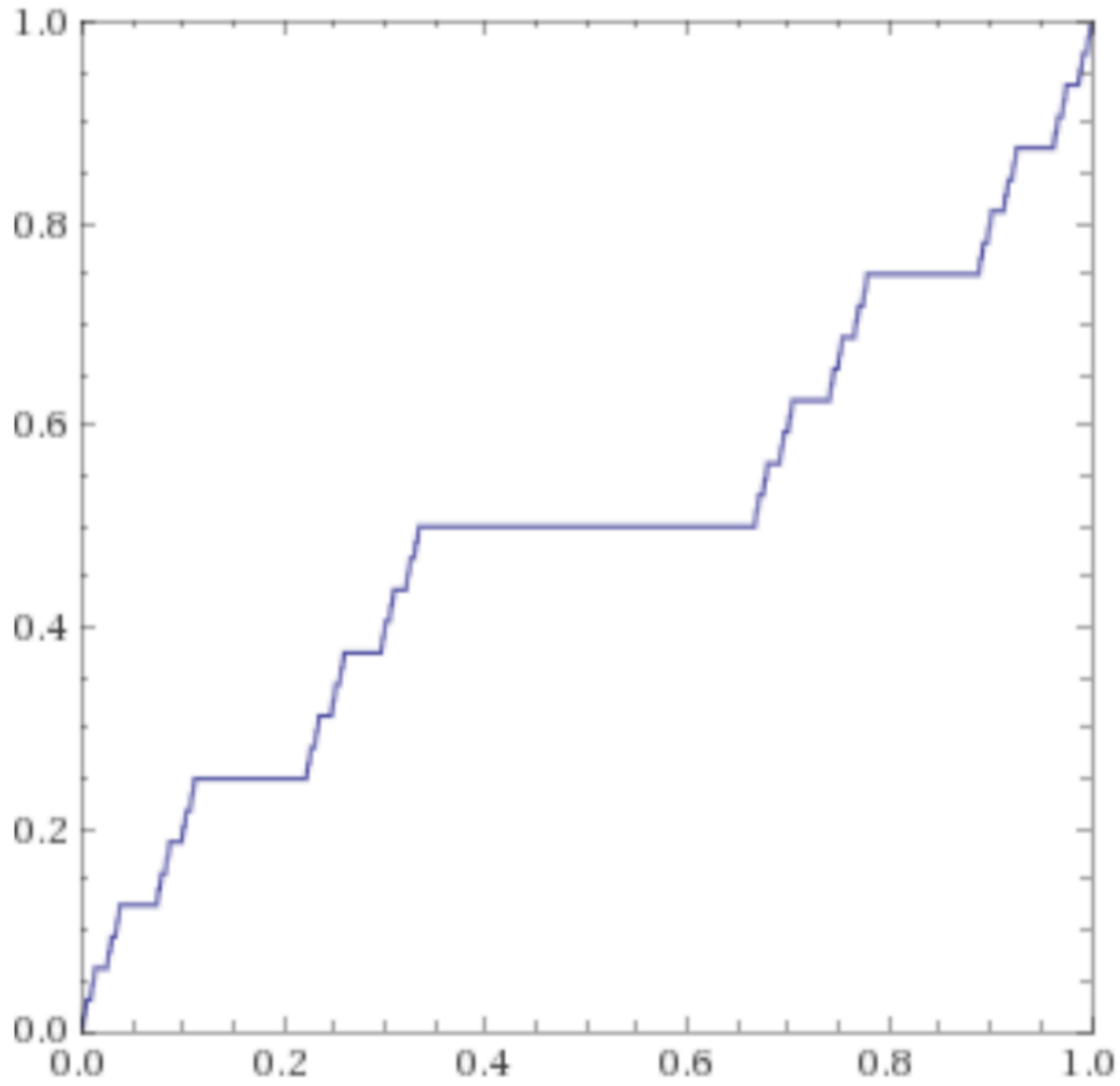



**Let** $\mu = \mu^{(f_1,f_2),\left(\frac{1}{2},\frac{1}{2}\right)}$ **such that**

$$\mu(A) = \sum_{i=1}^{N} \frac{1}{2}\mu(f_i^{-1}A) \text{ for every } A \in \mathscr{B}$$

# Example

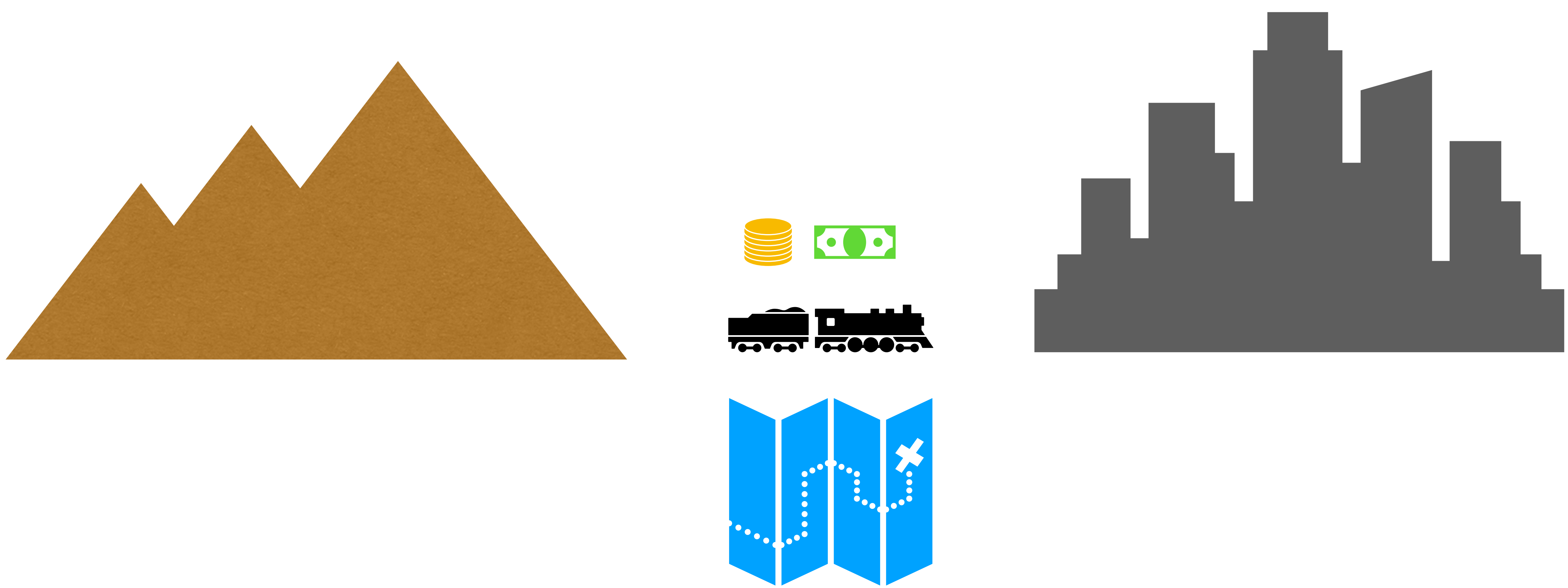
**The cumulative distribution function associated to this stationary probability measure corresponds to the Cantor function, that is** $f_{Cantor}(x) = \mu[0,x]$ .



# Monge's problem

**Is an engineering problem formulated in 1781:**

***Optimal transportation of the materials from a mine to another site.***



# Mathematical model for the Monge's problem

- A probability measure $\mu$ models the extracted mass.
- A probability measure $\nu$ models the constructed mass.
- A transport function $T$ models the initial and final position.
- A cost functions $c$ models the cost of transporting from one point to another.

# Monge's problem

**Given** $\mu$ **and** $\nu$ **, find** $T$ **such that** $\nu = \mu \circ T^{-1}$ **and such that minimises the total cost of transport**

$$\int c\,(x, Tx)\, d\mu(x)\,.$$

# Wasserstein Distance

- It corresponds to the transport problem when the cost function satisfies the axiom of a distance on a Polish space.
- Important in:
  - Statistics.
  - Limit theorems and approximation of probability measures.
  - Theory of propagation of chaos.
  - Boltzmann equations.
  - Mixing and convergence for Markov chains.
  - Rates of fluctuations of empirical measures.
  - Large-time behaviour of stochastic partial differential equations.
  - Hydrodynamic limits of systems of particles, Ricci curvature, Linearly rigid spaces, Towers of measures, Bernoulli automorphisms and classification of metric spaces, etc…

# Wasserstein Distance

- Given a Polish metric space $(\mathscr{X}, d)$ and $m \in [1,\infty)$ . For any two probability measures $\mu$ and $\nu$ on $\mathscr{X}$.
  The Wasserstein distance of order $m$ between $\mu$ and $\nu$ is defined by

$$W_m(\mu,\nu) := \inf \left\{ \left[\mathbb{E}d(X,Y)^m\right]^{\frac{1}{m}} : \ \textbf{law}(X) = \mu, \ \textbf{law}(Y) = \nu \right\}.$$

- When $m = 1$ it is called first Wasserstein distance, when $m = 2$ it is called second Wasserstein distance, etc…

# Kantorovich-Rubinstein duality theorem

**When** $\mathcal{X} = \mathbb{R}$

$$W_1(\mu, \nu) = \sup \left\{ \left| \int f d\mu - \int f d\nu \right| : \|f\|_{Lip} \leq 1 \right\}$$

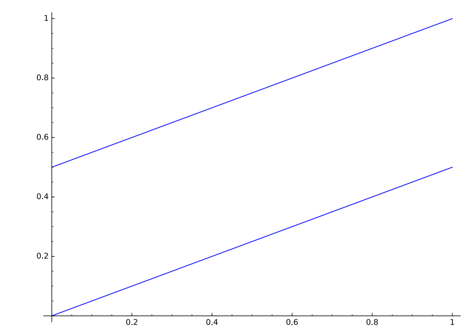
**This formula gives a reformulation of the first Wasserstein distance.**

# Jon Fraser's problem

- Let:
  - $\mathscr{X} = \mathbb{R}^d$.
  - An iterated function system $f = (f_1, f_2, \ldots, f_N)$.
  - Two probability vectors $p = (p_1, p_2, \ldots, p_N)$ and $q = (q_1, q_2, \ldots, q_N)$.
  - Two stationary probability measures $\mu = \mu^{(f,p)}$ **and** $\nu = \mu^{(f,q)}$ associated to $(f, p)$ **and** $(f, q)$, respectively.
- Find or estimate $W_m(\mu, \nu)$.

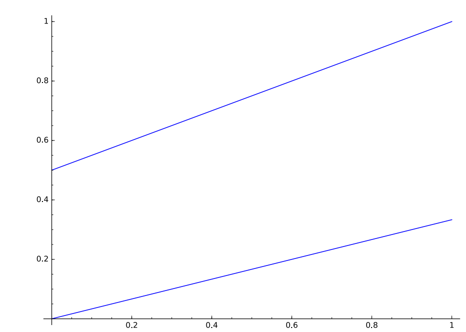
# Partial solutions

- **Theorem 1 [J. Fraser, 2015]** Explicit formula for the first Wasserstein distance in the case




$$\mathcal{X} = \mathbb{R}, f_1(x) = \rho x + t_1, f_2(x) = \rho x + t_2, \rho \in \left(0, \frac{1}{2}\right), 0 < \rho + t_1 < t_2 \leq 1.$$

- **Theorem 2 [J. Fraser, 2015]** Estimation for the second Wasserstein distance in the case




$$\mathcal{X} = \mathbb{R}, f_1(x) = \rho x + t_1, f_2(x) = \rho x + t_2, \rho \in \left(0, \frac{1}{2}\right), 0 < \rho + t_1 < t_2 \leq 1.$$

- **Theorem 3 [I.C and M. Pollicott, 2018]** Explicit formula for the first Wasserstein distance in the case

$$\mathcal{X} = \mathbb{R}, f_1(x) = \rho_1 x + t_1, f_2(x) = \rho_2 x + t_2, \rho_1, \rho_2 \in (0,1), 0 < \rho + t_1 \leq t_2 \leq 1.$$

# Main steps in proof Theorem 3

- **Lemma 1 [Dall'Aglio-Vallender].** Let $\mu$ and $\nu$ be probability measures on $\mathbb{R}$. Then

$$W_1(\mu,\nu) = \int_{-\infty}^{\infty} \left| F(t) - G(t) \right| dt,$$

where $F$ and $G$ are the cumulative distribution functions of $\mu$ and $\nu$, respectively.

- **Lemma 2 [A. Quas].** Suppose that $p \neq q$, then then function $D : [0,1] \to [0,1]$ defined by $D(x) := (\mu^{(f,p)} - \mu^{(f,q)})[0,x]$ does not change sign.

- **Remark.** $$\int_0^1 x d\mu^{(f,p)} = \frac{pt_1 + (1-p)t_2}{1 - (p\rho_1 + (1-p)\rho_2)}.$$

# Lemma 1

We really need a weaker result (key lemma).

- **Lemma 1' [J. Bochi].** Let $\mu$ **and** $\nu$ be probability measures on $[0,1]$. Then

$$W_1(\mu,\nu) = \int_0^1 \left( \int_0^x C_{\mu,\nu}(t)dt \right) d(\mu - \nu)(x),$$

where

$$C_{\mu,\nu}(x) := \begin{cases} 1 & \textbf{if } (\mu-\nu)[x,1] > 0, \\ -1 & \textbf{if } (\mu-\nu)[x,1] < 0. \end{cases}$$

# Proof

Proof. Suppose that $f$ with $\|f\|_{\text{Lip}} \leq 1$ realises the supremum in $d_{W_1}(\mu, \nu)$. Then $f(x) = \int_0^x g(x)dx$, where $g : [0,1] \to [-1,1]$ is an integrable function. By an application of Fubini's theorem we have

$$\begin{aligned}
\int_0^1 f(x)d\mu(x) - \int_0^1 f(x)d\nu(x) &= \int_0^1 f(x)d(\mu - \nu)(x) \\
&= \int_0^1 \int_0^x g(t)dtd(\mu - \nu)(x) \\
&= \int_0^1 \int_t^1 g(t)d(\mu - \nu)(x)dt \\
&= \int_0^1 g(t) \int_t^1 d(\mu - \nu)(x)dt \\
&= \int_0^1 g(t)(\mu - \nu)[t,1]dt.
\end{aligned}$$

Because of our assumption that $f$ realises the supremum in $d_{W_1}(\mu, \nu)$, we have that $g(x) = C_{\mu,\nu}(x)$. □

# Jon Fraser's list of specific problems [J. Fraser, 2015]

- **Overlaps:** Partially solved (work in progress).
- **Different contraction ratios:** Solved in [I.C and M. Pollicott, 2018].
- **Higher and non-integer moments:** Open.
- **Higher dimensions:** Open.
- **More than two maps:** Solved in [I.C preprint, 2018.]
- **Extension of the lower bound:** Solved in [I.C and M. Pollicott, 2018].

# More than two maps

- **Theorem 4 [I.C preprint, 2018].** Let $f = (f_1, f_2, \ldots, f_N)$ be an iterated function systems of <u>positive Lipschitz contractions</u> on the unit interval such that

$$f_i(0,1) \cap f_j(0,1) = \varnothing \text{ for all } i \neq j\,.$$

If $(p, q)$ is a pair of probability vectors in $(0,1)^N$ such that

$$\sum_{i=1}^{m} p_i - q_i \geq 0 \text{ ( or } \leq \text{) for every } m = 1,2,\ldots,N\,.$$

Then

$$W_1\left(\mu^{(f,p)}, \mu^{(f,q)}\right) = \left| \int_0^1 x d\mu^{(f,p)}(x) - \int_0^1 x d\mu^{(f,p)}(x) \right|\,.$$

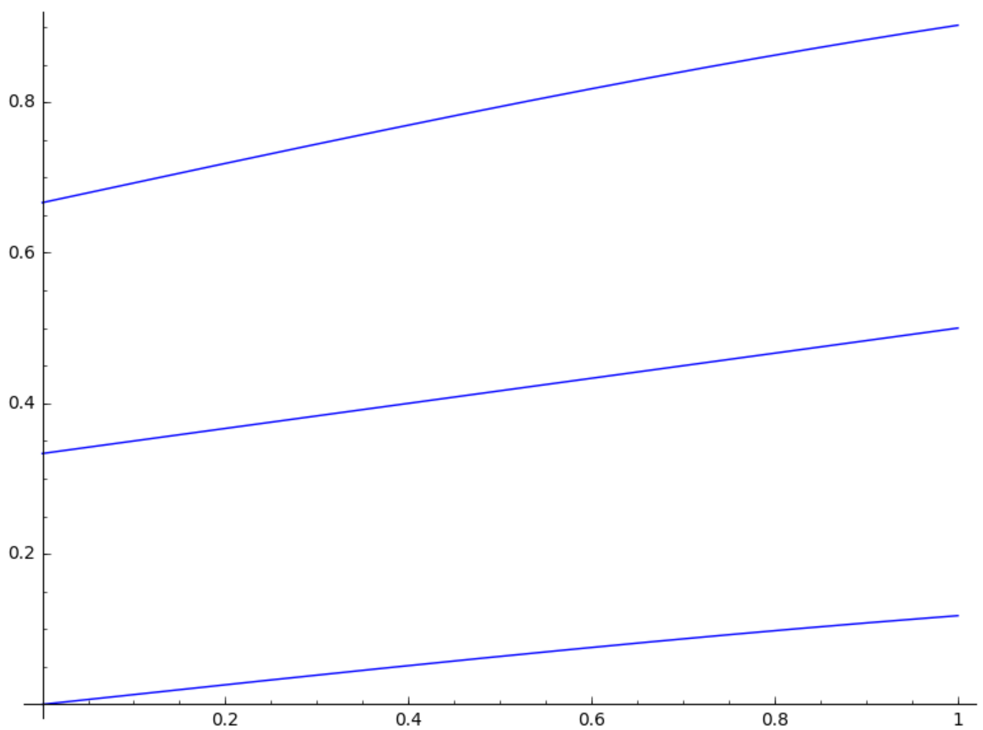
# Positive Lipschitz contractions

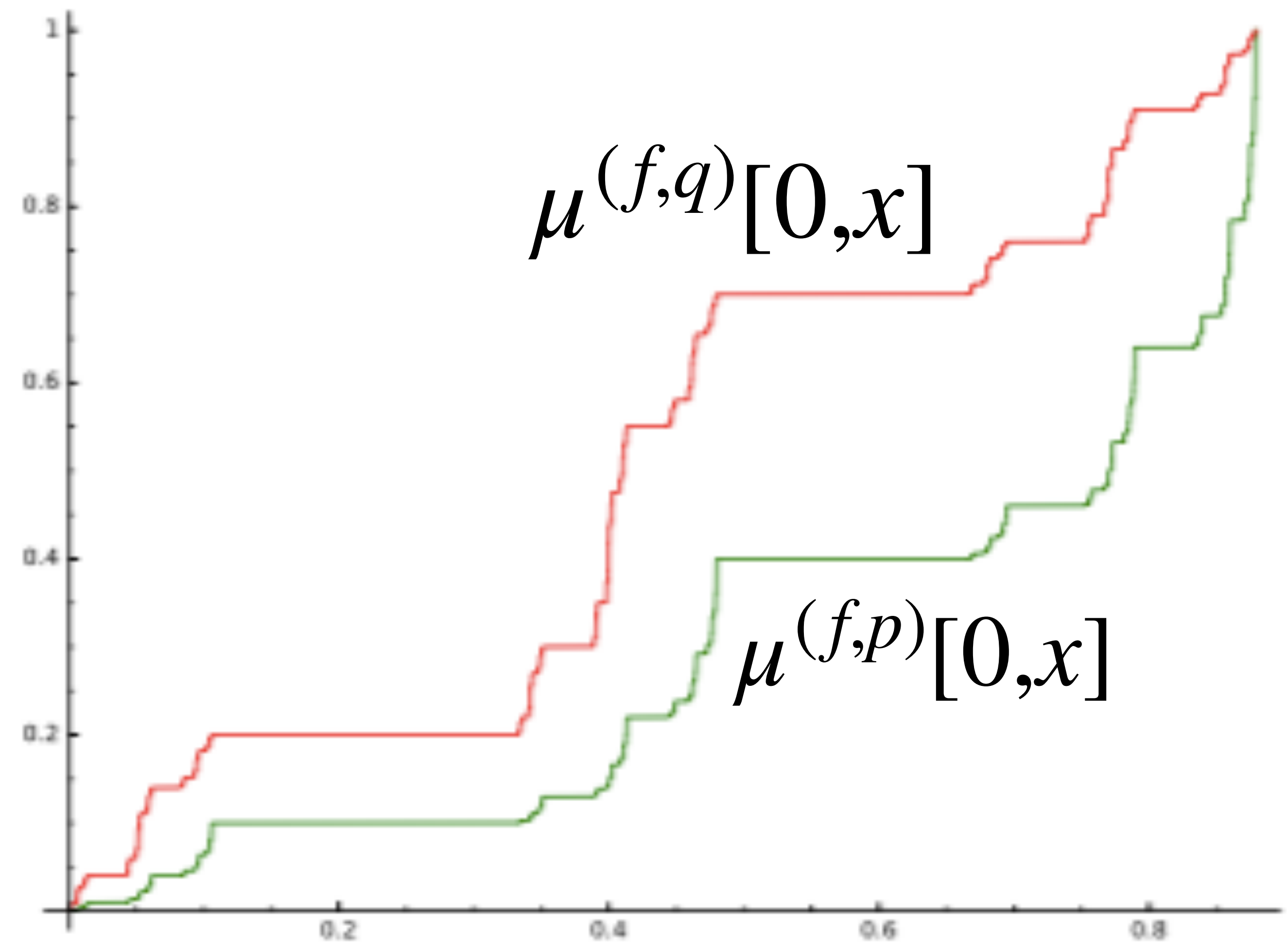
A Lipschitz contraction is a map $f : \mathbb{R} \to \mathbb{R}$ such that

$$\|f\|_{Lip} := \sup_{x,y\in\mathbb{R}} \frac{\left|f(x) - f(y)\right|}{|x - y|} < 1.$$

A positive Lipschitz contraction is a differentiable Lipschitz contraction map with positive derivative.

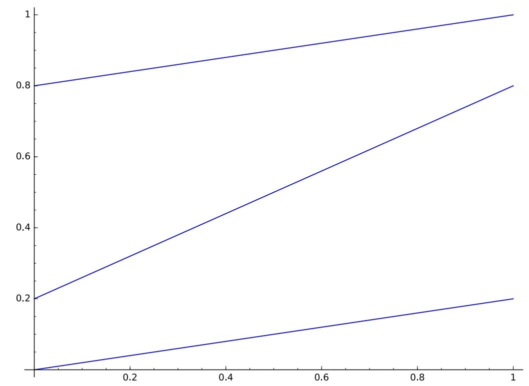
# Example more than two maps



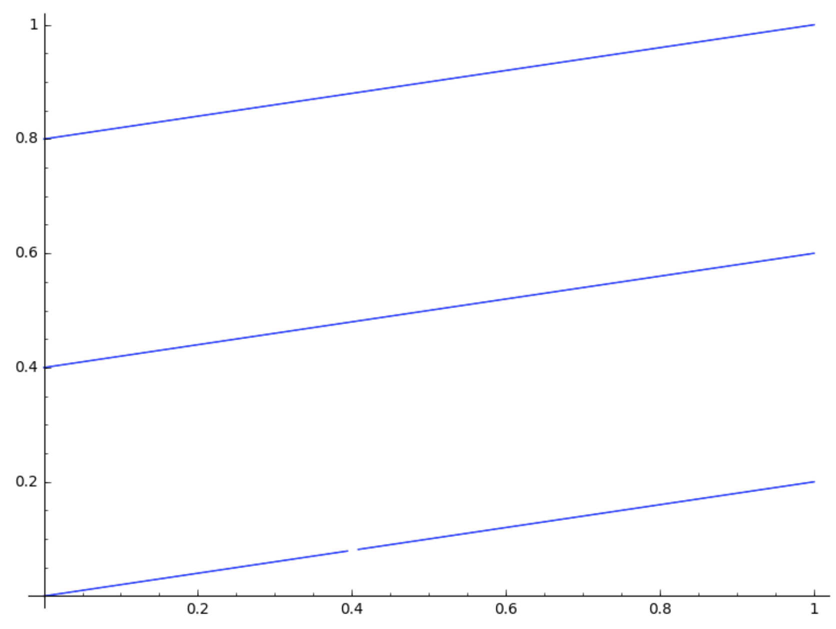

p = (0.1, 0.3, 0.6) and q = (0.2, 0.5, 0.3).

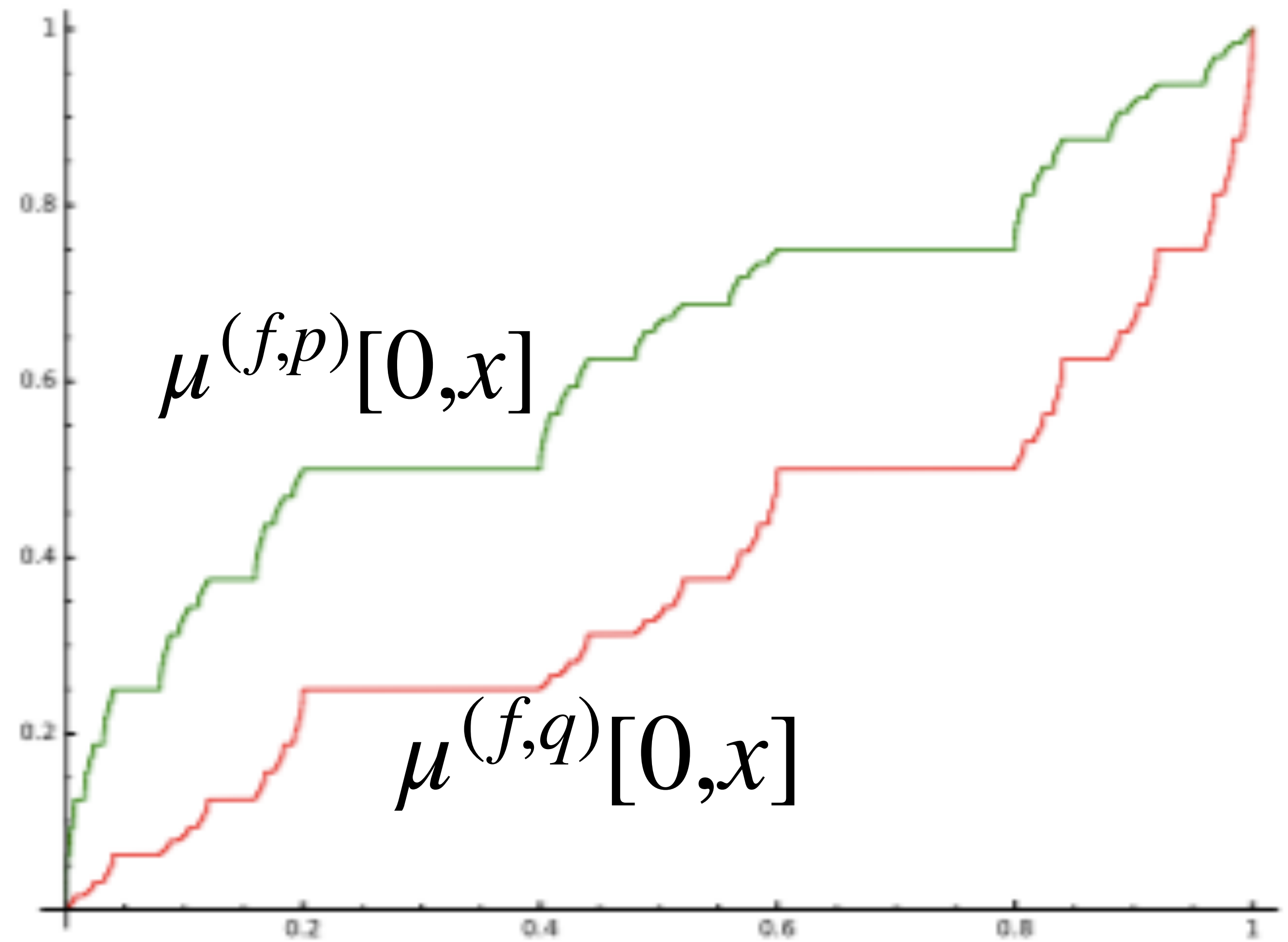
# More than two affine maps




- **Corollary.** Let $f_i : [0,1] \to [0,1]$ defined by $f_i(x) = \rho_i x + t_i$, where $\rho_i \in (0,1), t_i \in [0,1), \rho_i + t_i \leq t_{i+1}, i = 1,\ldots,N$.
  If $(p,q)$ is a pair of probability vectors in $(0,1)^N$ such that
  $$\sum_{i=1}^{m} p_i - q_i \geq 0 \text{ ( or } \leq \text{) for every } m = 1,2,\ldots,N.$$
  Then
  $$W_1\left(\mu^{(f,p)}, \mu^{(f,q)}\right) = \left| \frac{\sum_i p_i t_i}{1 - \sum_i p_i \rho_i} - \frac{\sum_i q_i t_i}{1 - \sum_i q_i \rho_i} \right|.$$

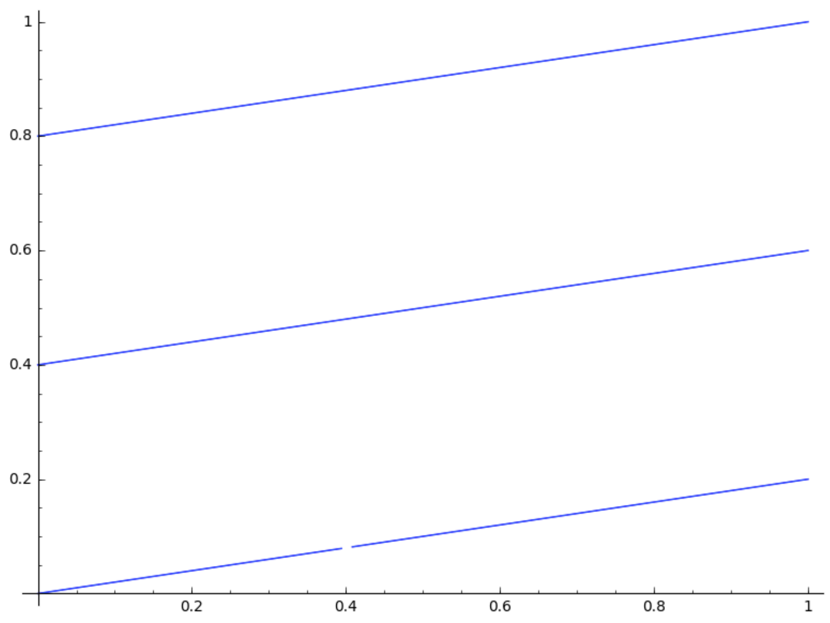
# Example more than two affine maps

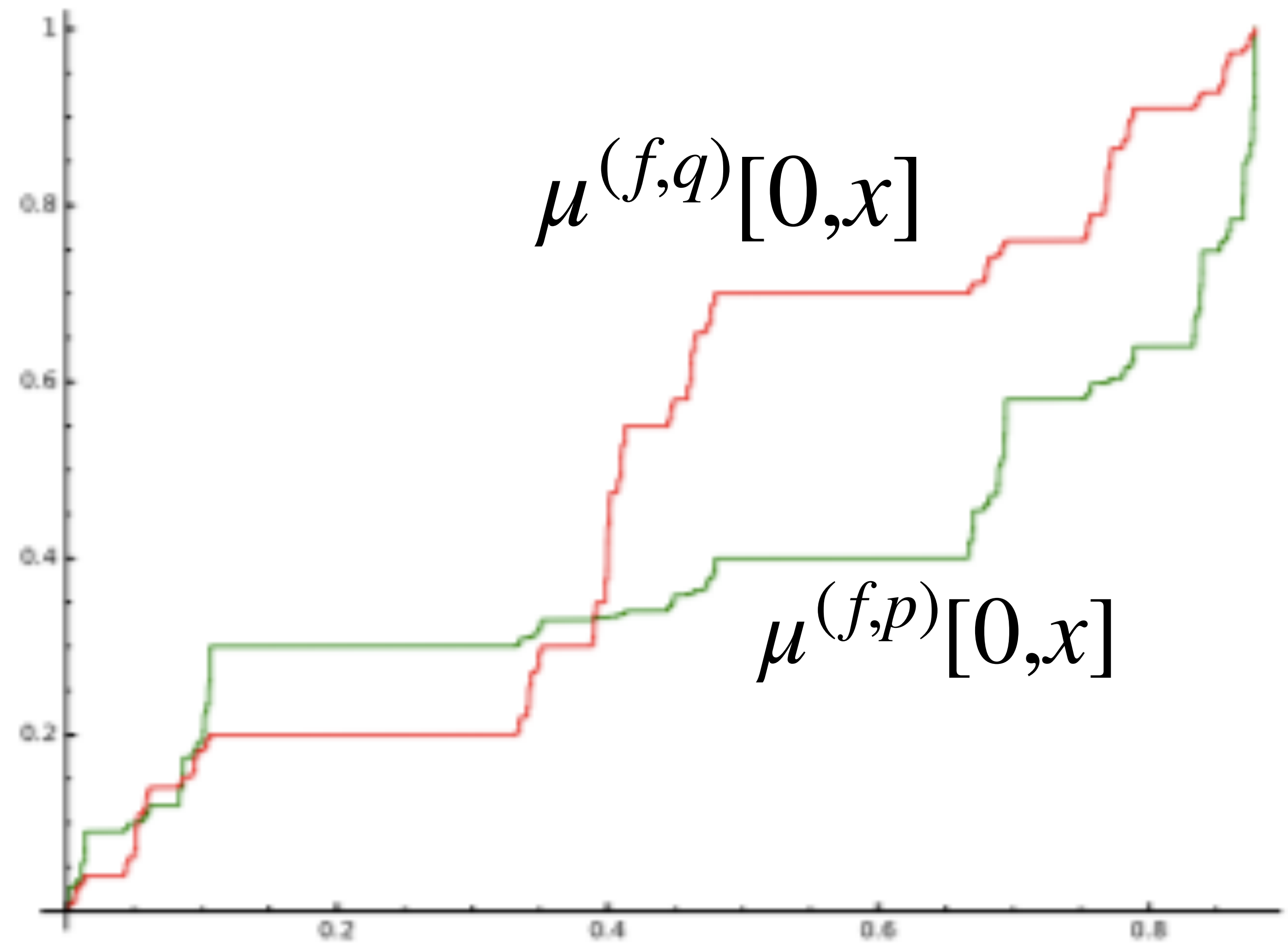



p = (1/2, 1/4, 1/4) and q = (1/4, 1/4, 1/2)

# Example more than two maps where the theorem does not apply




p = (0.3, 0.1, 0.6) and q = (0.2, 0.5, 0.3).

# Other cases?

- **Non-necessarily positive Lipchitz contractions?**
- $W_1\left(\mu^{(f,p)}, \mu^{(g,q)}\right)$ **when the iterated function systems** $f, g$ **are non necessarily the same and both contains only positive Lipschitz contraction?**
- $W_1\left(\mu^{(f,p)}, \mu^{(g,q)}\right)$ **when the iterated functions systems are non necessarily the same and** $f$ **contains only positive Lipschitz contractions whereas** $g$ **not?**

# Non-necessarily positive Lipchitz contractions

Let $r \in (2,\infty)$.

Let consider the iterated function system defined by $f_1(x) = \frac{x}{r},$

$$f_2(x) = 1 - \frac{x}{r}.$$

**Theorem [I.C preprint, 2018].** Let $k \in \mathbb{N}$ and $r \in (2k+1,\infty)$.

Then for $p = (p_1, p_2) = \left(\frac{1}{2k+1}, \frac{2k}{2k+1}\right)$ and $q = (p_2, p_1)$ we have that

$$W_1(\mu, \nu) = \int_0^1 c_r(x) d\left(\mu - \nu\right)(x)$$

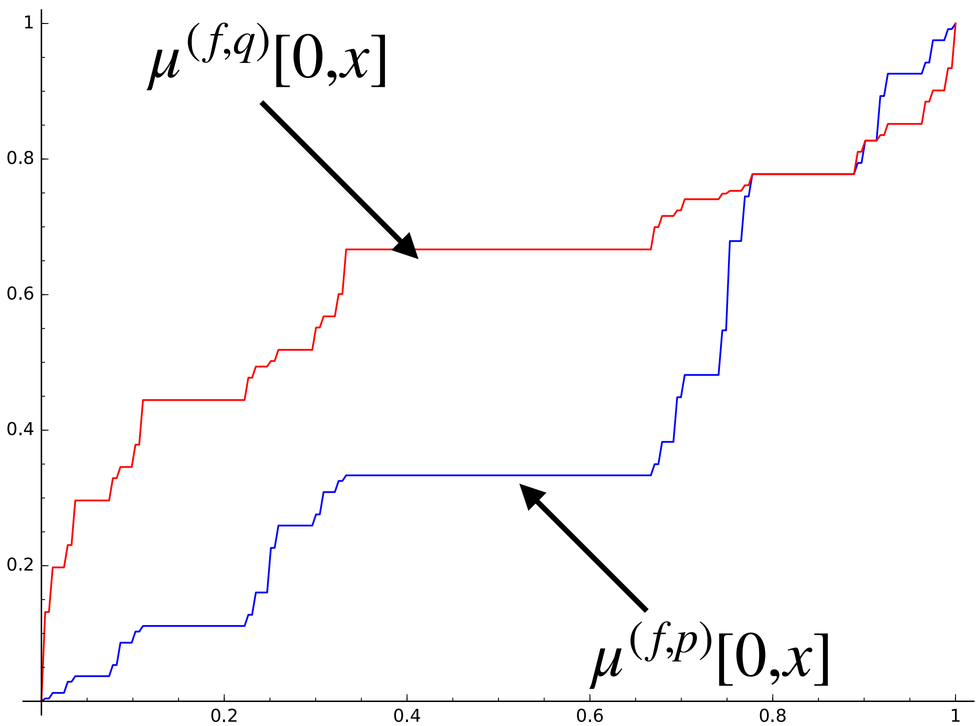
where

$$c_r(x) := \begin{cases} -x & \textbf{if } x < \frac{r^2}{r^2+1}, \\ x & \textbf{if } x > \frac{r^2}{r^2+1}. \end{cases}$$

# Example: Non-necessarily positive Lipchitz contractions

$$f_1(x) = \frac{x}{3}, \qquad p = \left(\frac{1}{3}, \frac{2}{3}\right),$$

$$f_2(x) = 1 - \frac{x}{3}. \qquad q = \left(\frac{2}{3}, \frac{1}{3}\right).$$



# Non-necessarily the same iterated function system

**Theorem [I.C preprint, 2018].** Let $f = (f_1, f_2)$ and $g = (g_1, g_2)$ be iterated function systems of positive Lipschitz contractions on the unit interval.

Suppose that $f_1(0) = g_1(0), f_2(0) = g_2(0), g_1(x) \leq f_1(x), g_2(x) \leq f_2(x)$ for all $x \in [0,1]$.

If $(p, q)$ is a pair of probability vectors $p = (p_1, 1 - p_1)$ and $q = (q_1, 1 - q_1)$ such that $p_1 \leq q_1$.

Then

$$W_1(\mu^{(f,p)}, \mu^{(g,q)}) = \int_0^1 x d\left(\mu^{(f,p)} - \mu^{(g,q)}\right)(x)$$

# Non-necessarily the same affine iterated function systems

**Corollary.** Let $f_i, g_i : [0,1] \to [0,1]$ be defined by $f_i(x) = \alpha_i x + t_i$, $g_i(x) = \beta_i x + t_i$

where $\rho_i \in (0,1), \beta_i \in (0,\rho_i]$ and $t_i \in (0,\rho_i]$ for $i = 1,2.$

Assume that $f = (f_1, f_2)$ and $g = (g_1, g_2)$ satisfy that

$$f_1(0,1) \cap f_2(0,1) = g_1(0,1) \cap g_2(0,1) = \varnothing .$$

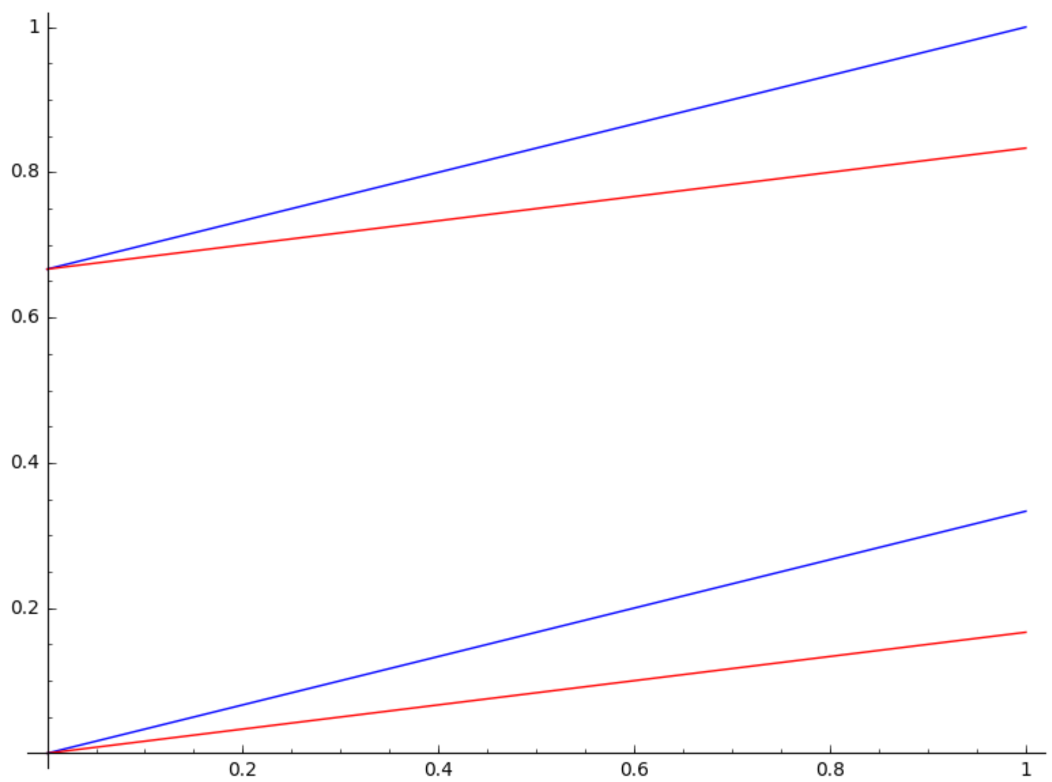
If $(p, q)$ is a pair of probability vectors $p = (p_1, 1 - p_1)$ and $q = (q_1, 1 - q_1)$ such that $p_1 \leq q_1$. Then

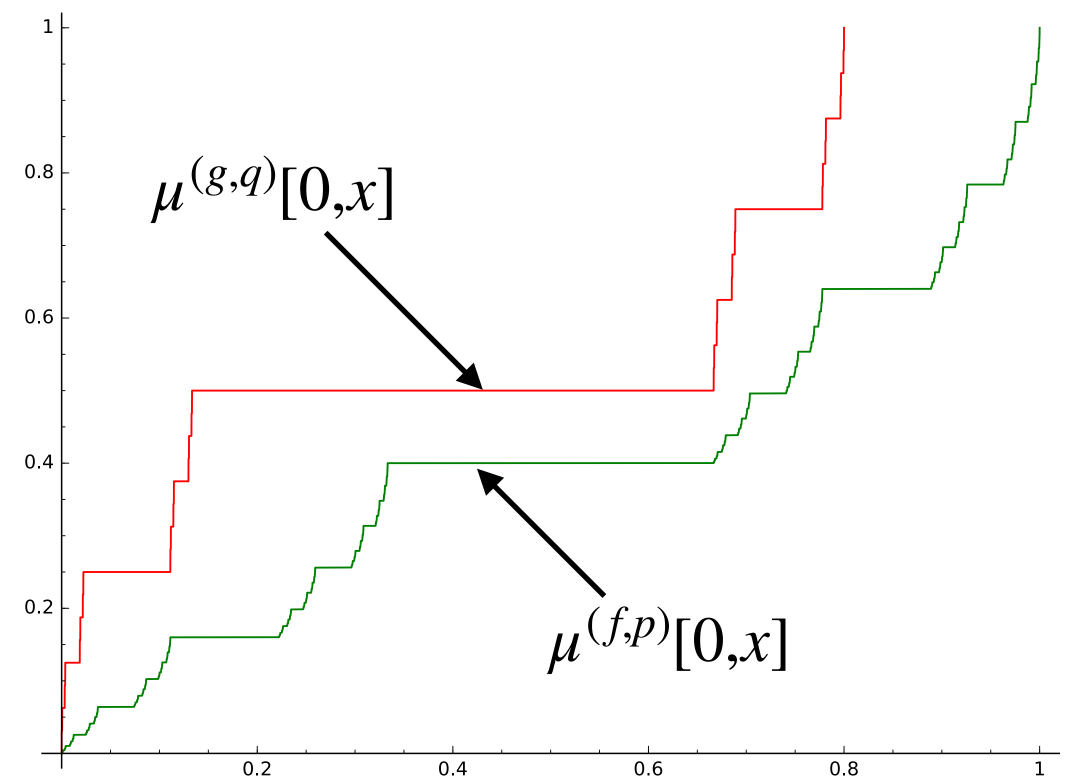
$$W_1(\mu^{(f,p)}, \nu^{(g,q)}) = \frac{\sum_i q_i t_i}{1 - \sum_i q_i \beta_i} - \frac{\sum_i p_i t_i}{1 - \sum_i p_i \alpha_i} .$$

# Example: Non-necessarily the same affine iterated function systems

$$f_1(x) = \frac{x}{3}, \qquad g_1(x) = \frac{x}{6}, \qquad p = \left(\frac{2}{5}, \frac{3}{5}\right),$$

$$f_2(x) = \frac{x}{3} + \frac{2}{3}. \qquad g_2(x) = \frac{x}{6} + \frac{2}{3}. \qquad q = \left(\frac{1}{2}, \frac{1}{2}\right).$$

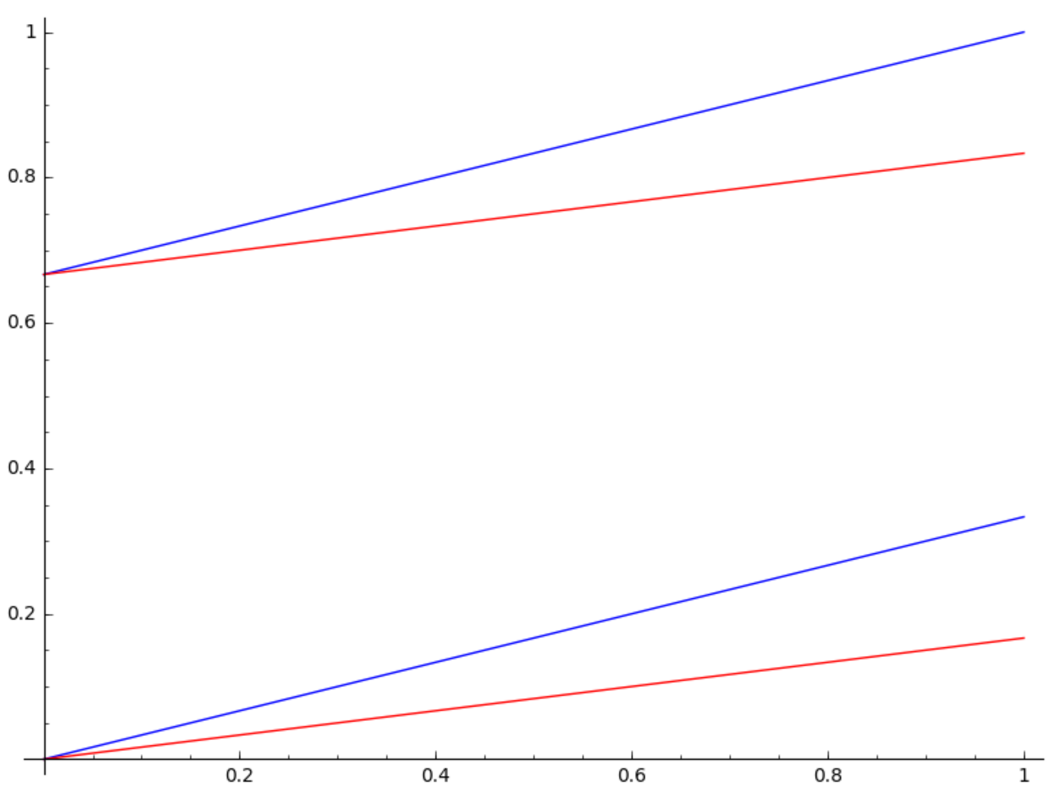


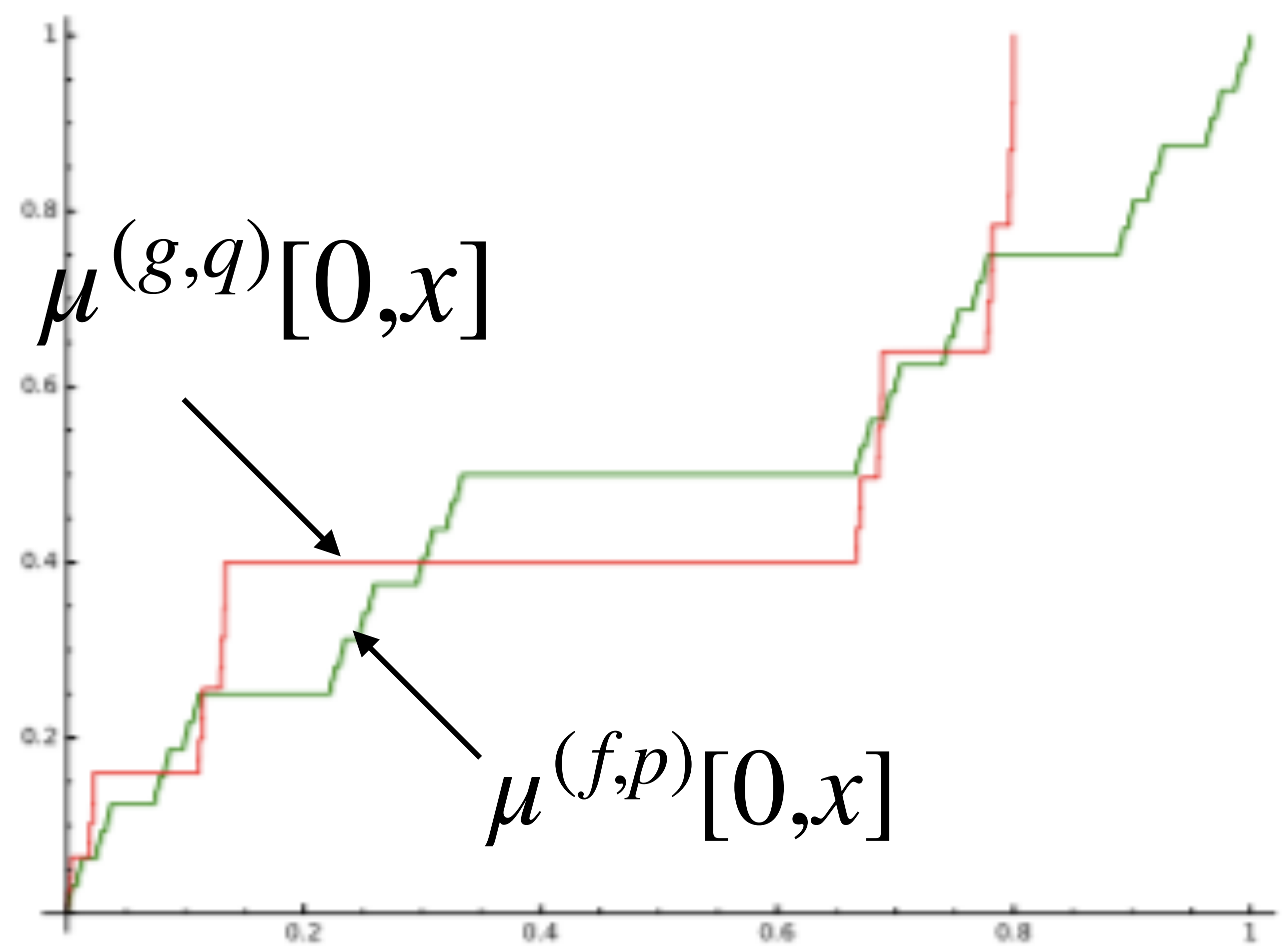



# Example: Non-necessarily the same affine iterated function systems, where theorem does not work

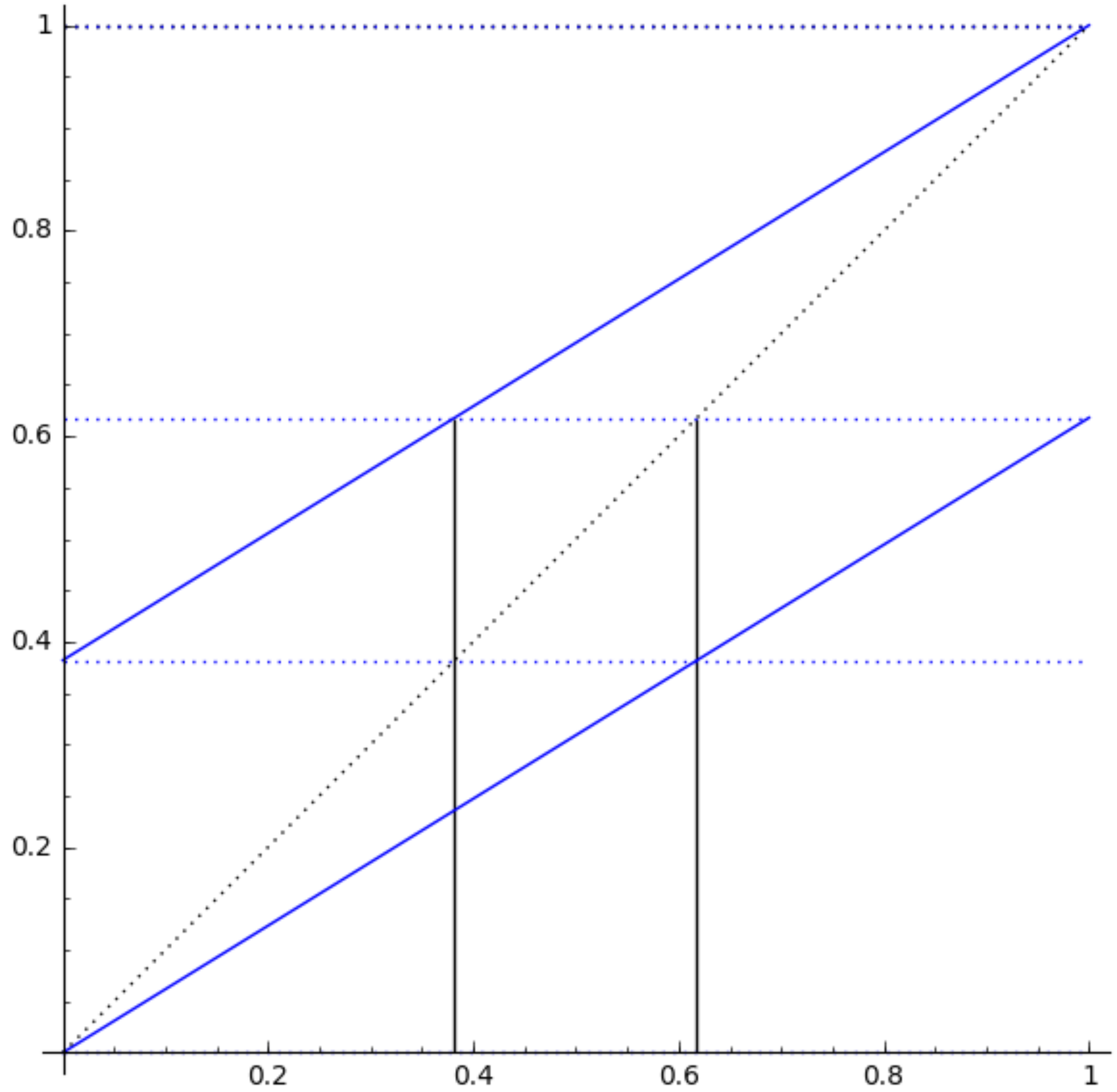
$$f_1(x) = \frac{x}{3}, \qquad g_1(x) = \frac{x}{6}, \qquad p = \left(\frac{1}{2}, \frac{1}{2}\right),$$

$$f_2(x) = \frac{x}{3} + \frac{2}{3}. \qquad g_2(x) = \frac{x}{6} + \frac{2}{3}. \qquad q = \left(\frac{2}{5}, \frac{3}{5}\right).$$



# Overlaps



# Bernoulli convolutions

- Consider the iterated function systems $f_1(x) = \rho x$ and $f_2(x) = \rho x + 1 - \rho$, where $\rho \in (1/2,1)$ is the reciprocal of a simple Pisot number, i.e., the inverse of the unique positive root of the polynomial $x^k - x^{k-1} - \ldots - x - 1 \quad (k = 2,3,\ldots)$.

- The stationary probability measure associated to the weight $(1/2,1/2)$ is called Bernoulli convolution with parameter $\rho$.

- The stationary probability measure associated to the weights $(x,1-x)$ for $x \in (0,1)\setminus\{1/2\}$ is called biased Bernoulli convolution.

- Erdős proved that the Bernoulli convolution with parameter $\rho$ reciprocal of a Pissot number is totally singular.

- Feng studied multifractal formalism and give an explicit formula for the local dimension of biased Bernoulli convolutions in the case of $\rho \in (1/2,1)$ is the reciprocal of a simple Pisot number.

# First Wasserstein distance between Bernoulli convolutions

**Theorem.** Let $\rho \in (1/2,1)$ be the reciprocal of a simple Pisot number and

$$f_1(x) = \rho x,$$
$$f_2(x) = \rho x + 1 - \rho\,.$$

If $p, q$ are two probability vectors in $(0,1)^2$. Then

$$W_1\left(\mu^{(f,p)}, \mu^{(f,q)}\right) = \left| \int_0^1 x d\mu^{(f,p)}(x) - \int_0^1 x d\mu^{(f,p)}(x) \right|.$$

# Thanks !