Mixed L2/Wasserstein Distances

J.-D. Benamou (POEMS - INRIA Rocquencourt)

Outline :

A - Basics on the L2 MKP and the TCMKP.
B - Relaxation and mixed L2/Wass distance.

2003 M2AN paper : Numerical Resolution of An "Unbalanced" Mass Transport Problem.

A - Basics on the L2 MKP and the TCMKP.

L^2 Monge Kantorovitch / Mass transport Problem

Given two balanced ($\int \rho_0 = \int \rho_1 = 1$) mass densities $\rho_0,\ \rho_1 : R^d \mapsto R^+$.

MKP2 : Compute

$$d_{wass}(\rho_0, \rho_1)^2 = \inf_{M \in \mathcal{M}} \int_{R^d} |x - M(x)|^2 \rho_0(x) dx$$

$$\mathcal{M} = \{M : R^d \mapsto R^d,\ \rho_1(M)\, det(\nabla M) = \rho_0\}$$

Difficulties :
Constraint is NL. - $\mathcal{M}$ is a (very) large set.

Theorem : There exists a unique $M_{opt} = \nabla\psi$
$\psi\, :\, R^d \mapsto R$ convex.
(Brenier, ... See Villani's book).

Corollary : Weak form of Monge Ampere Equation (Caffarelli)

$$\rho_1(\nabla\psi)\, det(H\psi) = \rho_0$$

with non standard B.C. $\nabla\psi(supp(\rho_0)) = supp(\rho_1)$

Interesting FD Numerical methods for degenerate MA solution
(Oberman, Oberman/Froese).

Time Continuous MKP (Brenier B. ,2000) Add a Time dimension : $t \in [0,1]$

$$d_{wass}(\rho_0,\rho_1)^2 = \inf_{(\rho,v)\in\mathcal{T}} \int_0^1 \int \frac{1}{2}\rho(t,x)\,|v(t,x)|^2\,dx\,dt$$

$$\mathcal{T} = \{ \quad (\rho,v) : [0,1]\times R^d \mapsto R^+ \times R^d, \; s.t.$$

$$\rho(0,.) = \rho_0, \; \rho(1,.) = \rho 1 \; and$$

$$\partial_t \rho + \nabla.(\rho v) = 0 \; \}$$

Proof based on : $v_{opt}(t,x) = M_{opt}(x) - x.$

Why ? A. -TCMKP leads to converging and easy to implement Numerical method. : Ch. of Variables : $(\rho,v) \to (\rho, m := \rho\, v)$

$$d_{wass}(\rho_0,\rho_1)^2 = \inf_{(\rho,m)\in\mathcal{T}} \int_0^1 \int \frac{|m(t,x)|^2}{2\,\rho(t,x)}\,dx\,dt$$

Then remark that (for $\rho > 0$)

$$\frac{|m(t,x)|^2}{2\rho(t,x)} = \sup_{\{a,b\}\in K} [a(t,x)\rho(t,x) + b(t,x).m(t,x)]$$

$$K = \{\{a,b\} : R \times T^d \to R \times R^d, \;\; s.\, t. \;\; a + \frac{|b|^2}{2} \leq 0\,\}.$$

The constraints become

$$\mathcal{T} = \{ \;\; (\rho, m) : [0,1] \times R^d \mapsto R^+ \times R^d, \; s.t.$$

$$\rho(0,.) = \rho_0, \; \rho(1,.) = \rho_1 \; and$$

$$\nabla_{t,x} \cdot (\rho, m) = 0 \;\; \}$$

Augmented Lagrangian method (ALG2, Fortin-Glowinski) works fine, even for vanishing and discontinuous densities !

See Guittet 2003 SINUM paper for theory convergence ...

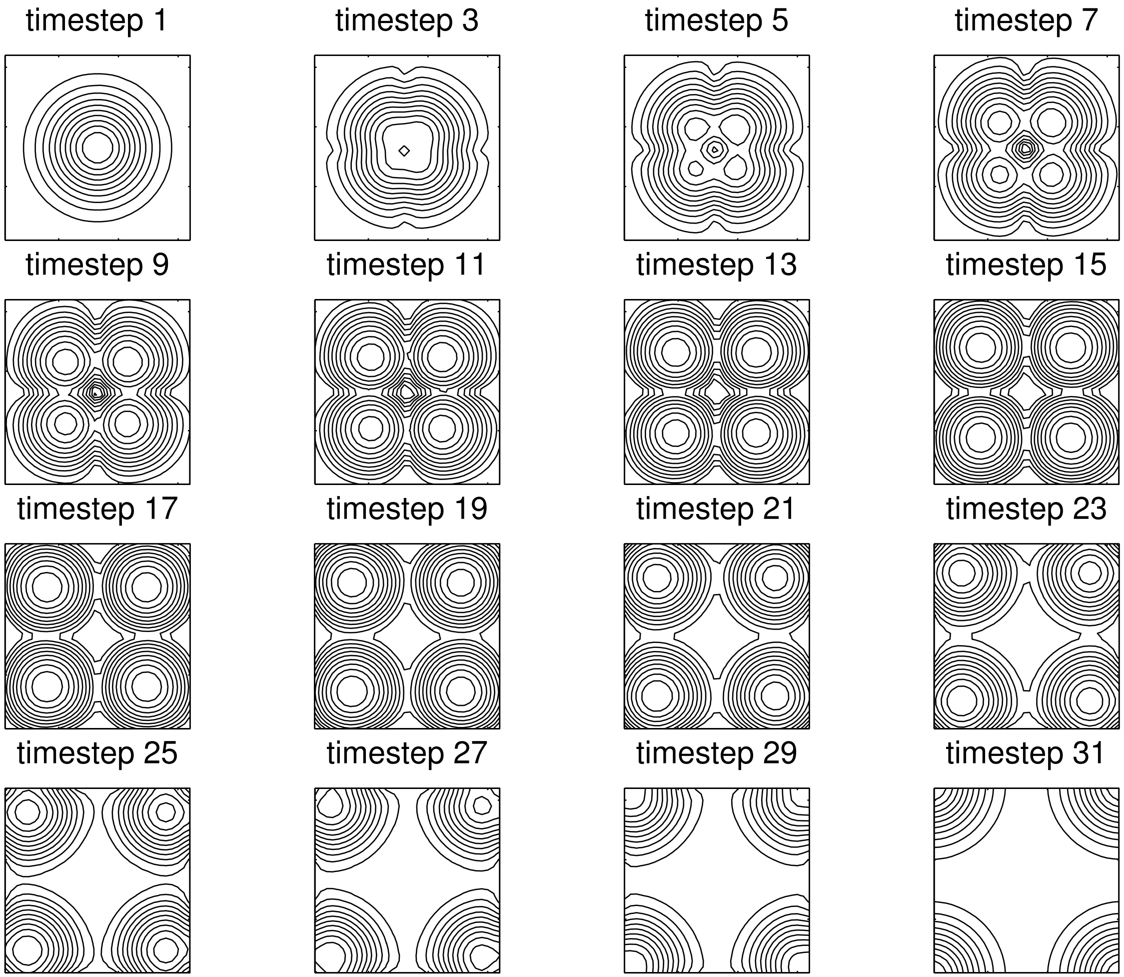
TCMKP is a MFG

Introduce a Lagrange multiplier $\phi(t,x)$ for constrains in $\mathcal{T}$. Optimality Conditions for the saddle point problem are :

$$\begin{cases} \partial_t \phi + \frac{|\nabla_x \phi|^2}{2} = 0, \\ \partial_t \rho + \nabla_x \cdot m = 0, \\ m = \rho \, \nabla \phi, \\ \rho(0,.) = \rho_0, \;\; \rho(1,.) = \rho_1. \end{cases}$$

$\rightarrow$ Example in Achdou, Camilli, Capuzzo-Dolcetta (2010).

Example : shift of a periodic array of Gaussians Level curves of $\rho(t^k, .)$



B - Relaxation of final density and mixed L2/Wass distance

Relaxation of final density

$$d_{mix}(\rho_0, \rho_1)^2 = \inf_{m \in \mathcal{T}_{relax}} J_{relax}(m)$$

$$J_{relax}(m) = \int_0^1 \int \frac{|m(t,x)|^2}{2\,\rho(t,x)}\,dx\,dt + \frac{\gamma}{2}\int |\rho(1,x) - \rho_1(x)|^2 dx$$

$(\gamma > 0)$

$$\mathcal{T}_{relax} = \{ \quad m : [0,1] \times T^d \mapsto \times T^d, \ s.t.$$

$$\rho(0,.) = \rho_0 \ and$$

$$\partial_t \rho + \nabla_x \cdot m = 0 \quad \}$$

Why ?

A. 1 - No constraints on ρ_1

A. 2 - "Classical" Optimal Control of a system governed by a PDE $\rightarrow$ Gradient methods apply

Gradient from Direct/Adjoint problem

$$J'_{relax}(m) = \frac{m}{\rho} - \nabla_x \phi$$

with

$$\begin{cases} \partial_t \phi + \frac{|m|^2}{2\,\rho^2} = 0, \\ \partial_t \rho + \nabla_x \cdot m = 0, \\ \rho(0,.) = \rho_0, \;\; \phi(1,.) = \gamma\,(\rho(1,.) - \rho_1) \end{cases}$$

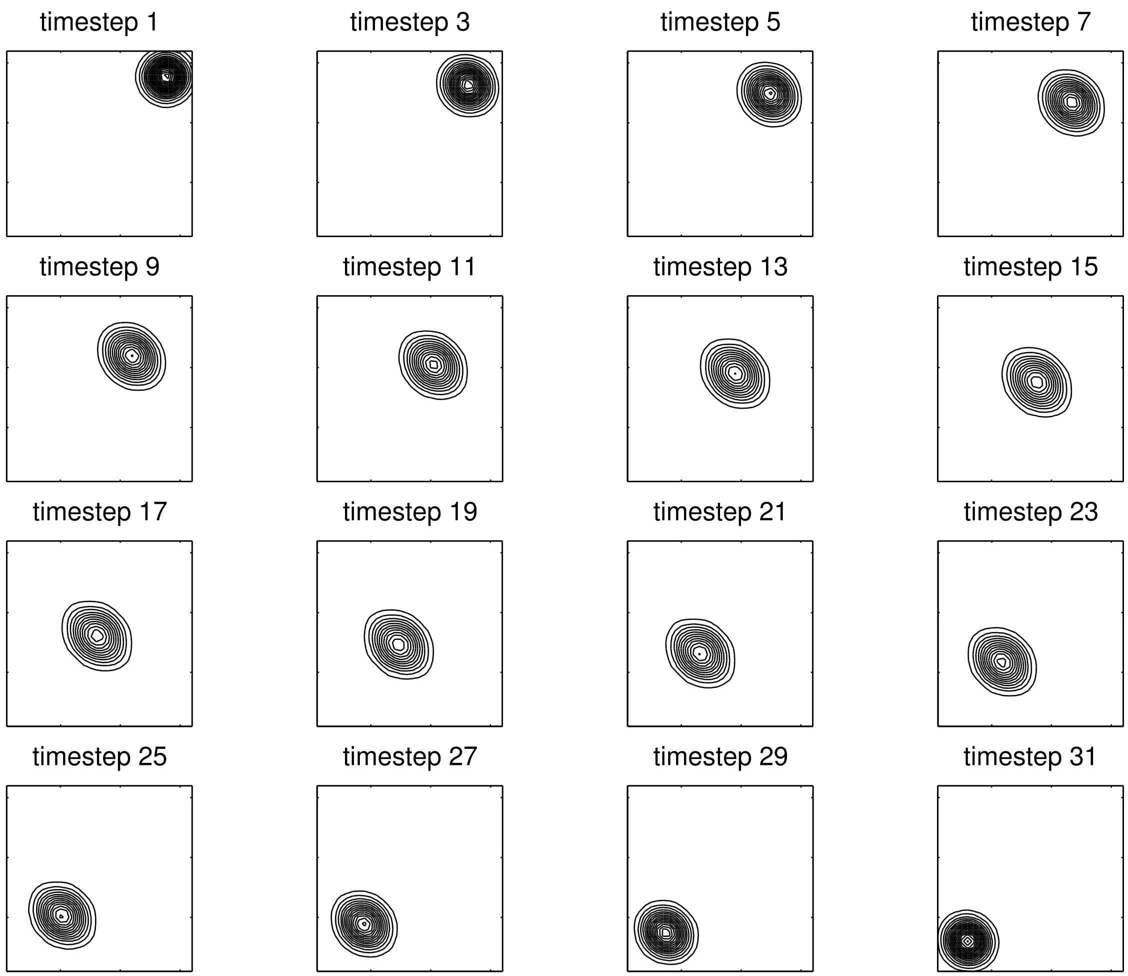
Problem : Only works when $\rho > \epsilon > 0$

Fix :

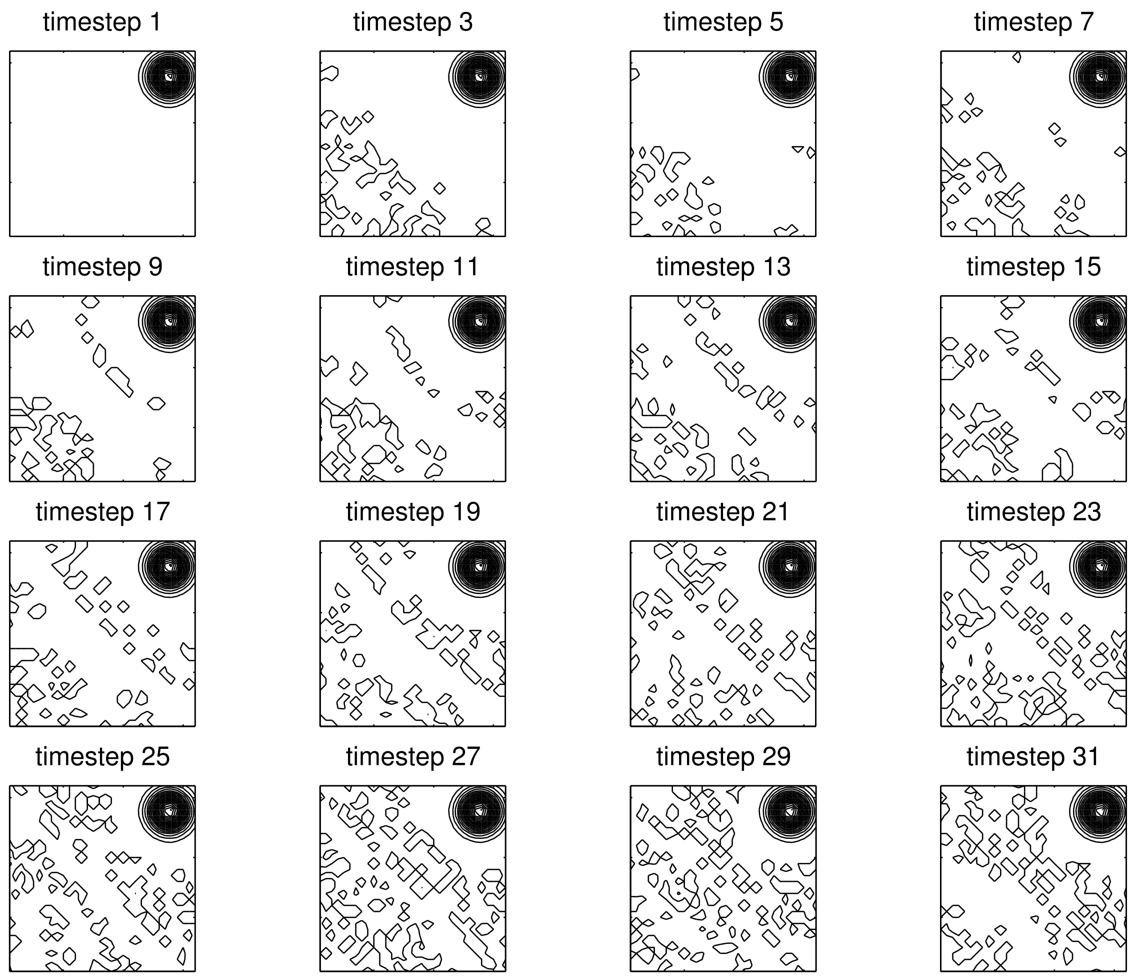
$$\frac{|\tilde{\rho}_1 - \rho_1|^2}{2} = \sup_{c \in R} [c(\tilde{\rho}_1 - \rho_1) - \frac{c^2}{2}]$$

(pointwise) and again apply ALG2.

Level curves of $\rho(t^k, .)$ for $\gamma = 100$ (non periodic, balanced data)



Level curves of $\rho(t^k, .)$ for $\gamma = 0.01$ (non periodic)



The "mixed distance"

$$d_{mix}(\rho_0, \rho_1)^2 = \inf_{\substack{\tilde{\rho}_1 \\ \tilde{\rho}_1 > 0 \\ \int \tilde{\rho}_1 dx = 1}} \{d_{wass}(\rho_0, \tilde{\rho}_1)^2 + \frac{\gamma}{2} d_{L^2}(\tilde{\rho}_1, \rho_1)^2\}$$

Why ?

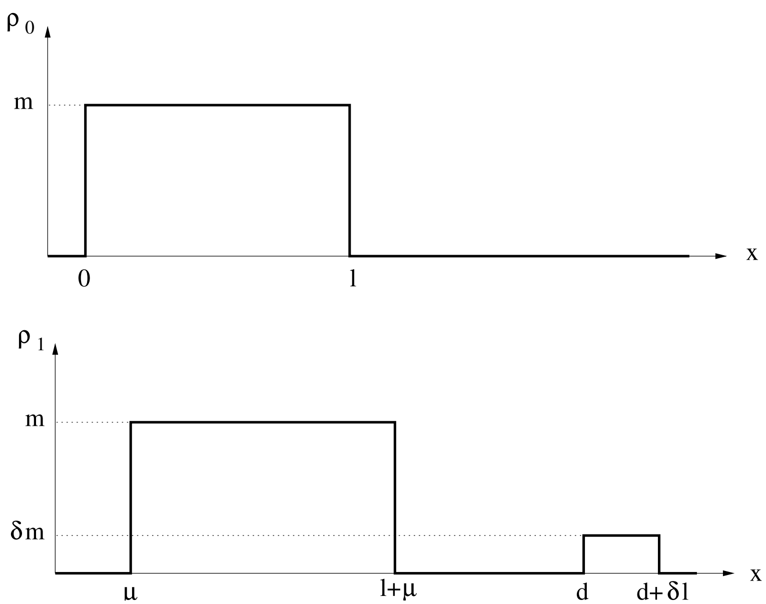
A. 1 - Wassertein distance is good at measuring drift errors ($d_{was}(\rho_0(.), \rho_0(. + \mu)^2 = |\mu|^2 \int \rho_0 dx$).

A. 2 - Data is rarely balanced.

A. 3 - Estimate separately drift and "other" errors.

Idealized example (suggested by M. Cullen)

$$\rho_1(.) = \rho_0(. + \mu) + \delta\rho(.)$$



Goal : Find γ such that

$$d_{l^2}(\rho_1, \tilde{\rho_1}^{opt}_{\gamma})^2 = \int |\delta\rho|^2 \, dx$$

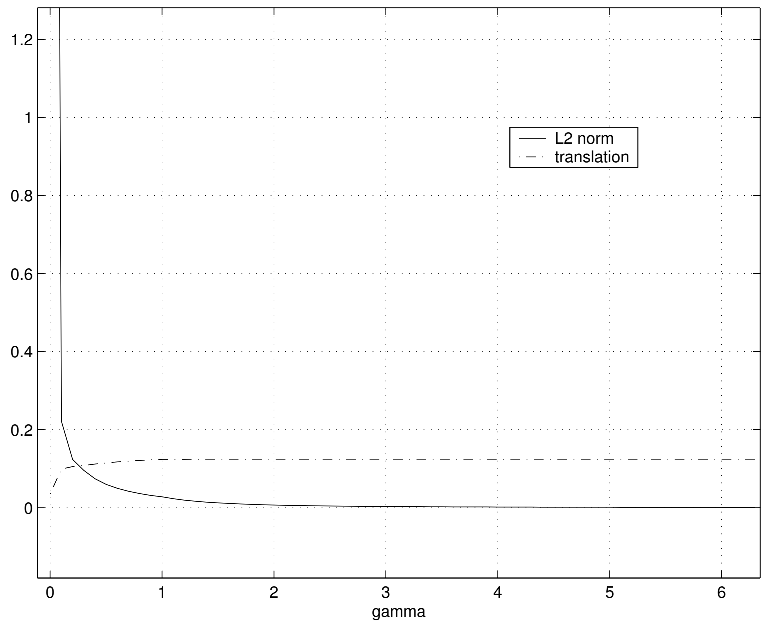
$$d_{wass}(\rho_1, \tilde{\rho_1}^{opt}_{\gamma})^2 = \mu^2 \int \rho_0 \, dx$$

Remark

$$\gamma \to d_{l^2}(\rho_1, \tilde{\rho_1}_\gamma^{opt})^2$$

$$\gamma \to d_{wass}(\rho_1, \tilde{\rho_1}_\gamma^{opt})^2$$

are monotone



Proposed algorithm

- Guess $\delta\rho$.

For ex: $\delta\rho_G = \rho_1 - \rho_0$ outside supp(ρ_0) expanded by the possible drift error.

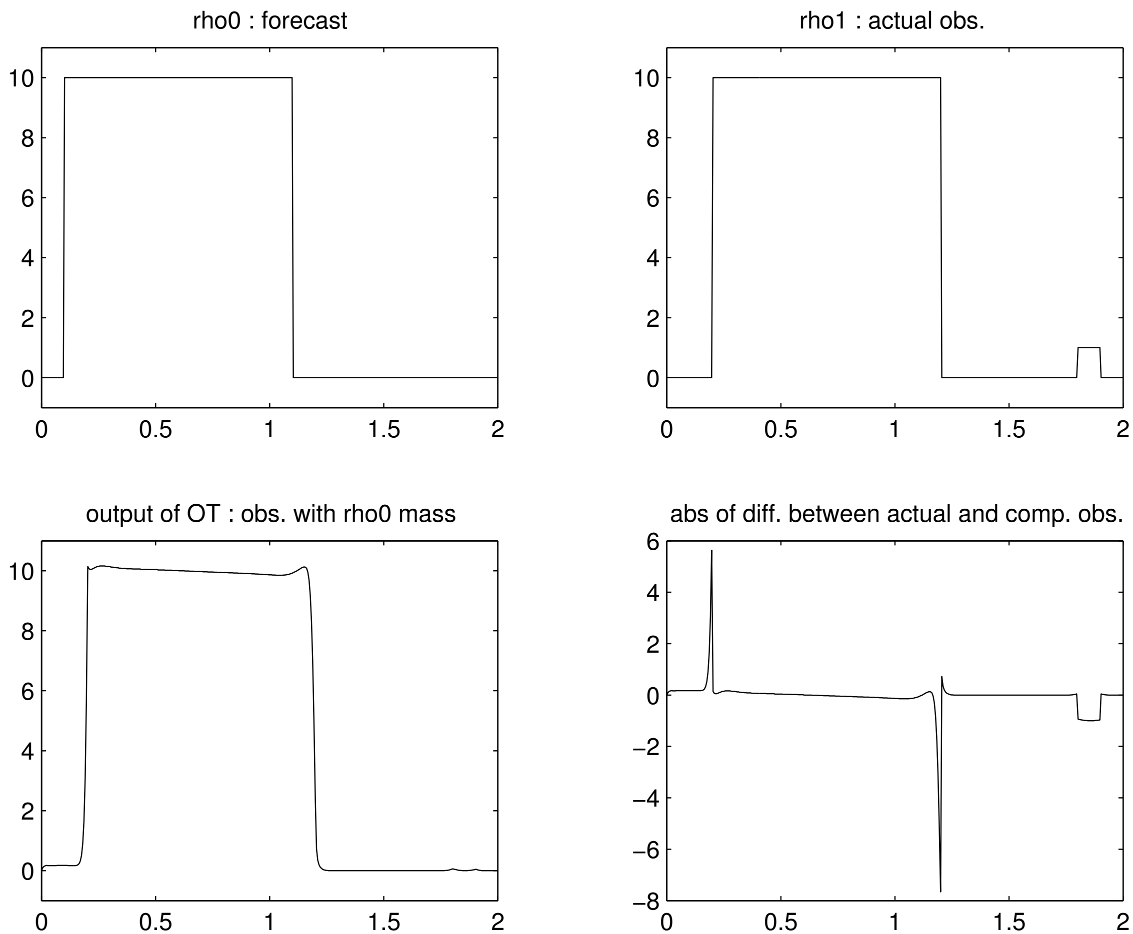
- Find γ s.t. $d_{l^2}(\rho_1, \tilde{\rho_1}_\gamma^{opt})^2 = \int |\delta\rho_G|^2\, dx$.

Num. Results

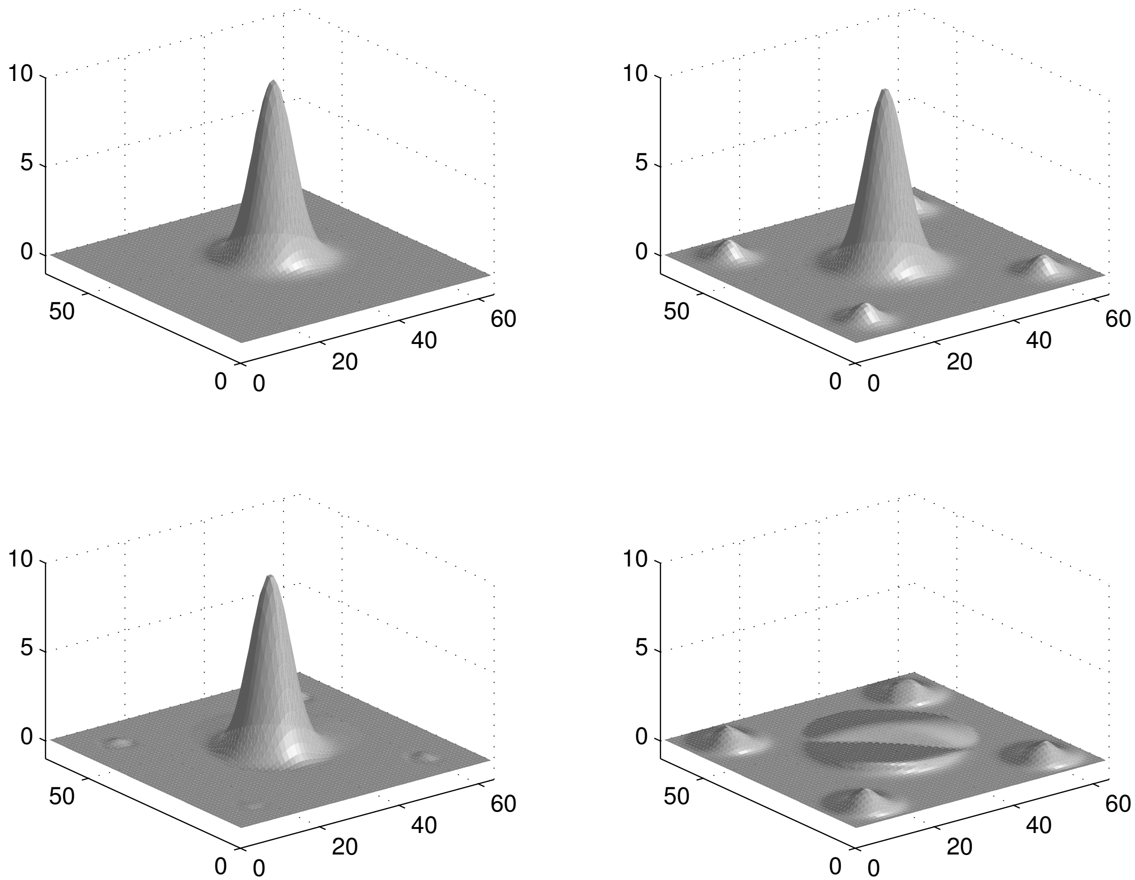
$$\gamma = 0.2783, \;\; d_{L^2}(\rho_1, \tilde{\rho_1})^2 = 0.0989, \;\; \mu = \sqrt{\frac{d_{was}(\rho_0, \tilde{\rho_1})^2}{\int \rho_0(x)\, dx}} = 0.1080$$

(exact values 0.1).

Row wise : ρ_0, ρ_1, $\tilde{\rho_1}$, $\tilde{\rho_1} - \rho_1$.



Exact :$\int |\delta\rho(x)|^2 dx \simeq 0.078539774 \quad |\mu| = 0.1\sqrt{2}$



Estimated :$\gamma = 0.2159 \quad d_{l^2}(\rho_1, \tilde{\rho_1})^2 = 0.0784689113 \quad |\mu| = 0.13.$

THANK YOU !