Unnormalized Optimal Transport

Stanley Osher

UCLA, 2019

Goals

- A simple and natural way to compare densities with unnormalized/unbalanced total mass.

W. Gangbo

W. Li

S. Osher

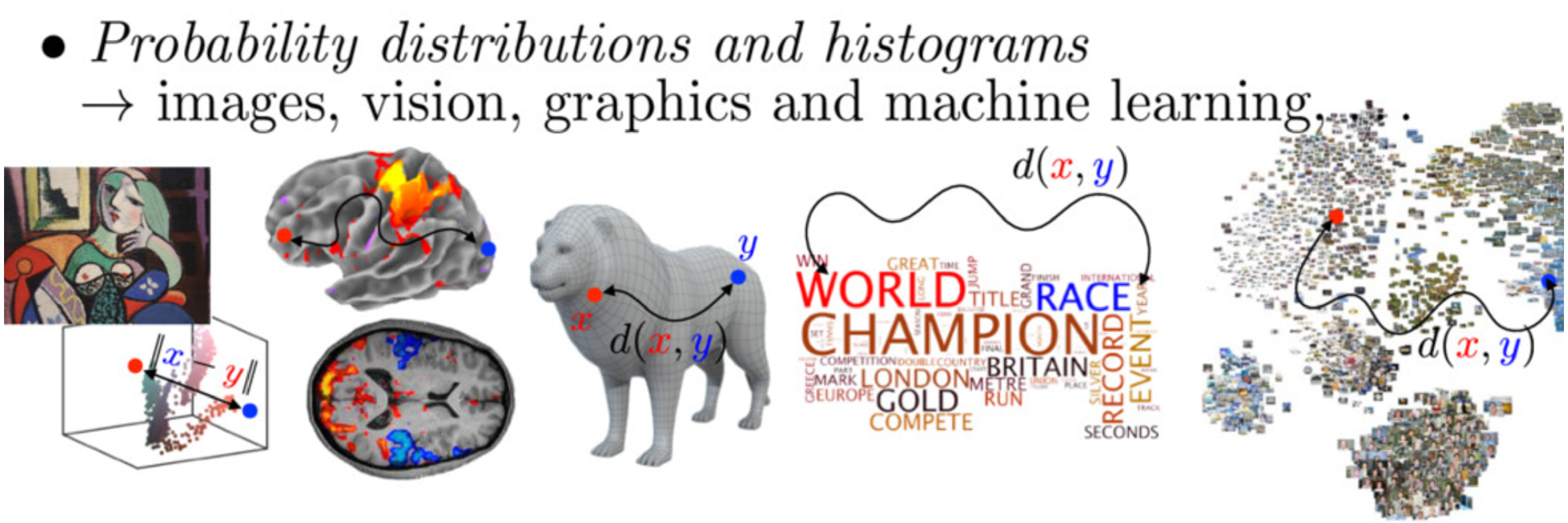
M. Puthawala

GLOP

Distance among histograms

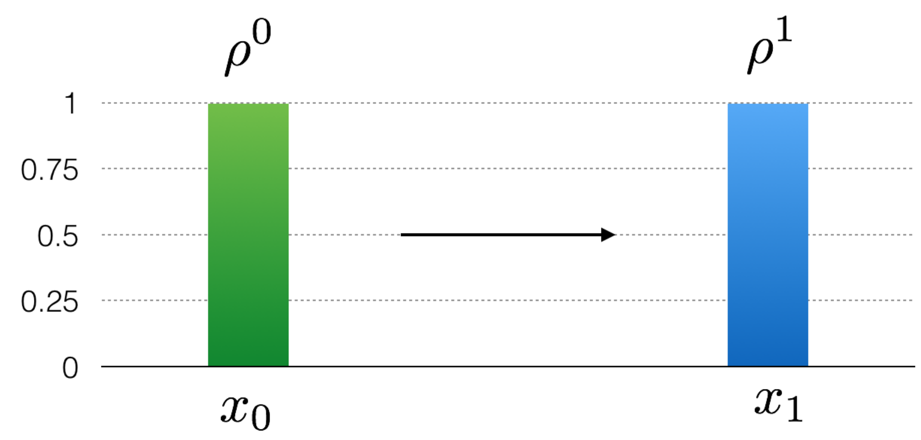
Measuring the closeness among density functions (histograms) plays crucial roles in applications, such as

- Image processing and Inverse problems (Li et. al 2018, Yang et.al 2018, Puthawala et.al. 2018);
- Machine learning (Lin et. al 2018);
- Mean field games (Chow et. al 2018).



Transport Distance

Optimal transport provides a particular distance (W) among histograms, which relies on the distance on sample spaces (ground cost c).



Denote $X_0 \sim \rho^0 = \delta_{x_0}$, $X_1 \sim \rho^1 = \delta_{x_1}$. Compare

$$W(\rho^0, \rho^1) = \inf_{\pi \in \Pi(\rho^0, \rho^1)} \int \int c(x, y) \pi(x, y) dx dy = c(x_0, x_1);$$

Vs

$$\mathrm{TV}(\rho^0, \rho^1) = \int_\Omega |\rho^0(x) - \rho^1(x)| dx = 2;$$

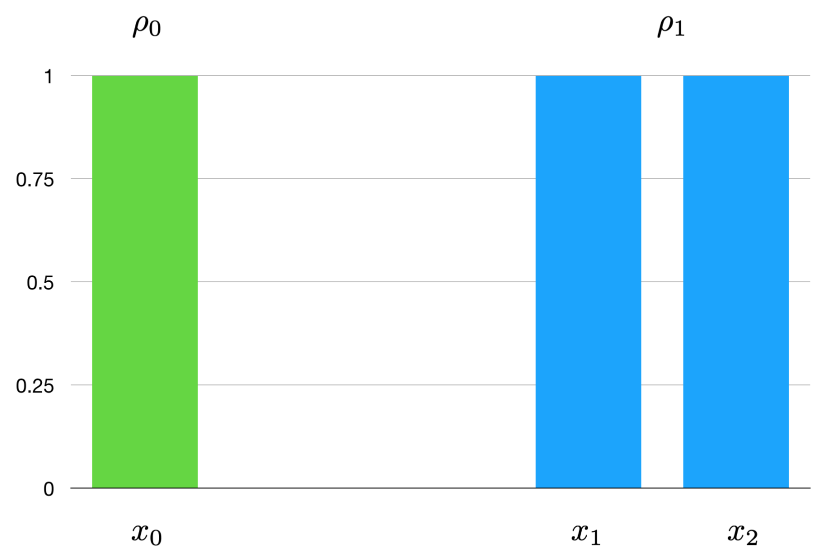
Vs

$$\mathrm{KL}(\rho^0 \| \rho^1) = \int_\Omega \rho^0(x) \log \frac{\rho^0(x)}{\rho^1(x)} dx = \infty.$$

Goals: Unnormalized Transport

Main Questions

In real applications such as inverse problems and image processing, one needs to measure unnormalized/unbalanced densities.



Solutions:

We propose a simple and natural modification of optimal transport to compare unnormalized/unbalanced densities, and introduce an efficient numerical scheme.

Related studies

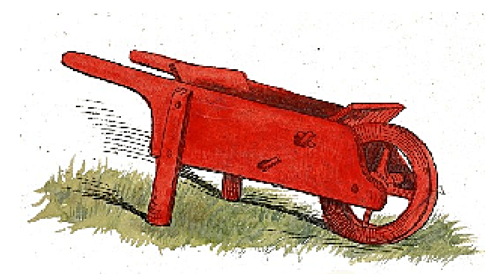
- Wasserstein-Fisher-Rao metric, L. Chizat, G. Peyre, B. Schmitzer, and F.-X. Vialard, *Journal of Functional analysis*.
- Hellinger-Kantorovich metric, M. Liero, A. Mielke, and G. Savare, *Inventiones mathematicae*.
- Free boundaries in optimal transport and Monge-Ampere obstacle problems, L. Caffarelli and R. McCann, *Annals of Mathematics*.
- Transport and equilibrium in non-conservative systems, L. Chayes and H. K. Lei, *Advances in Differential Equations* .
- Transport based image morphing with intensity modulation, J. Maas, M. Rumpf and S. Simon, *SSVM*.

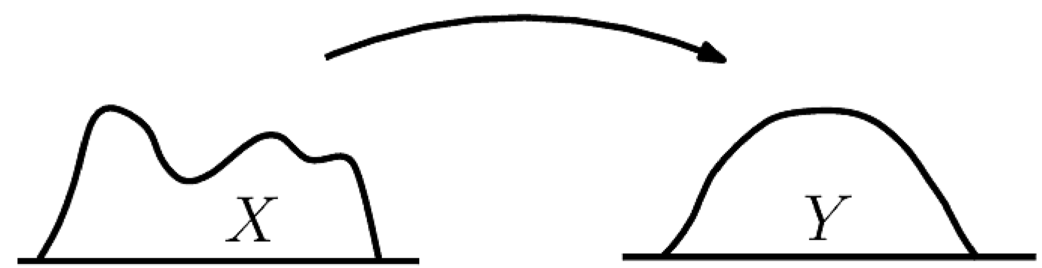
Compared to the above approaches, unnormalized OT has a closed-form

Unnormalized Monge-Ampere equation,

is able to be solved by a very simple and efficient *Primal-Dual algorithm* (Chambolle-Pock).

Optimal transport





What is the optimal way to move or transport the mountain with shape X, density $\rho^0(x)$ to another shape Y with density $\rho^1(y)$?

The optimal transport problem was first introduced by Monge in 1781, relaxed by Kantorovich by 1940. It introduces a particular metric on probability set. In literatures, the problem is often named Earth Mover's distance, Monge-Kantorovich problem and Wasserstein metric, etc.

Normalized Optimal Transport

Balanced case

$$\int_{x\in A} \rho_1(x)dx = \int_{T(x)\in A} \rho_0(x)dx,$$

where T is a smooth one to one map on $\mathbb{R}^d$:

$$\det(\nabla T(x))\rho_1(T(x)) = \rho_0(x),$$

This is called the Jacobian equation underdetermined.

L^p Monge–Kantorovich–Wasserstein distance

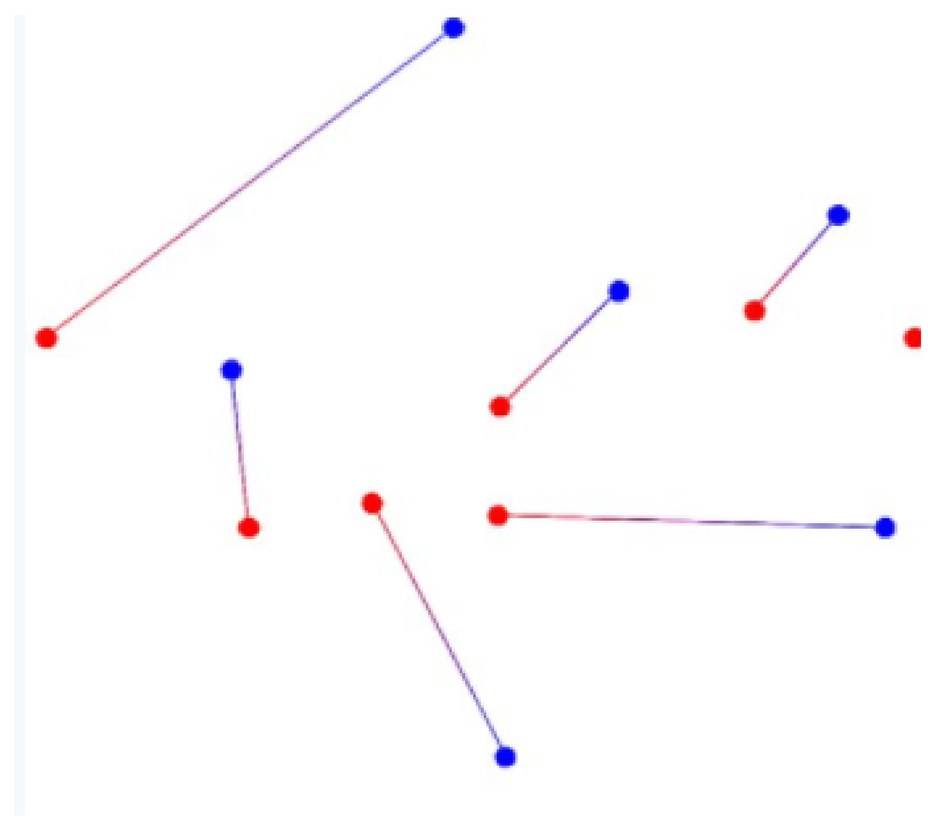
Given two measures ρ^0, ρ^1 with equal mass. Consider

$$(W_p(\rho_0, \rho_1))^p = \inf_T \quad \int_\Omega \|x - T(x)\|^p \rho_0(x) dx$$

where the infimum is among all transport maps T, which transfers $\rho^0(x)$ to $\rho^1(x)$, i.e.

$$\rho_0(x) = \rho_1(T(x)) \det(\nabla T(x)) \ .$$

The minimizer T is the optimal transfer, which solves L^p Monge-Kantorovich problem.



Monge–Ampere equation

Brenier showed for $p = 2$, uniqueness of optimal transfer T, such that

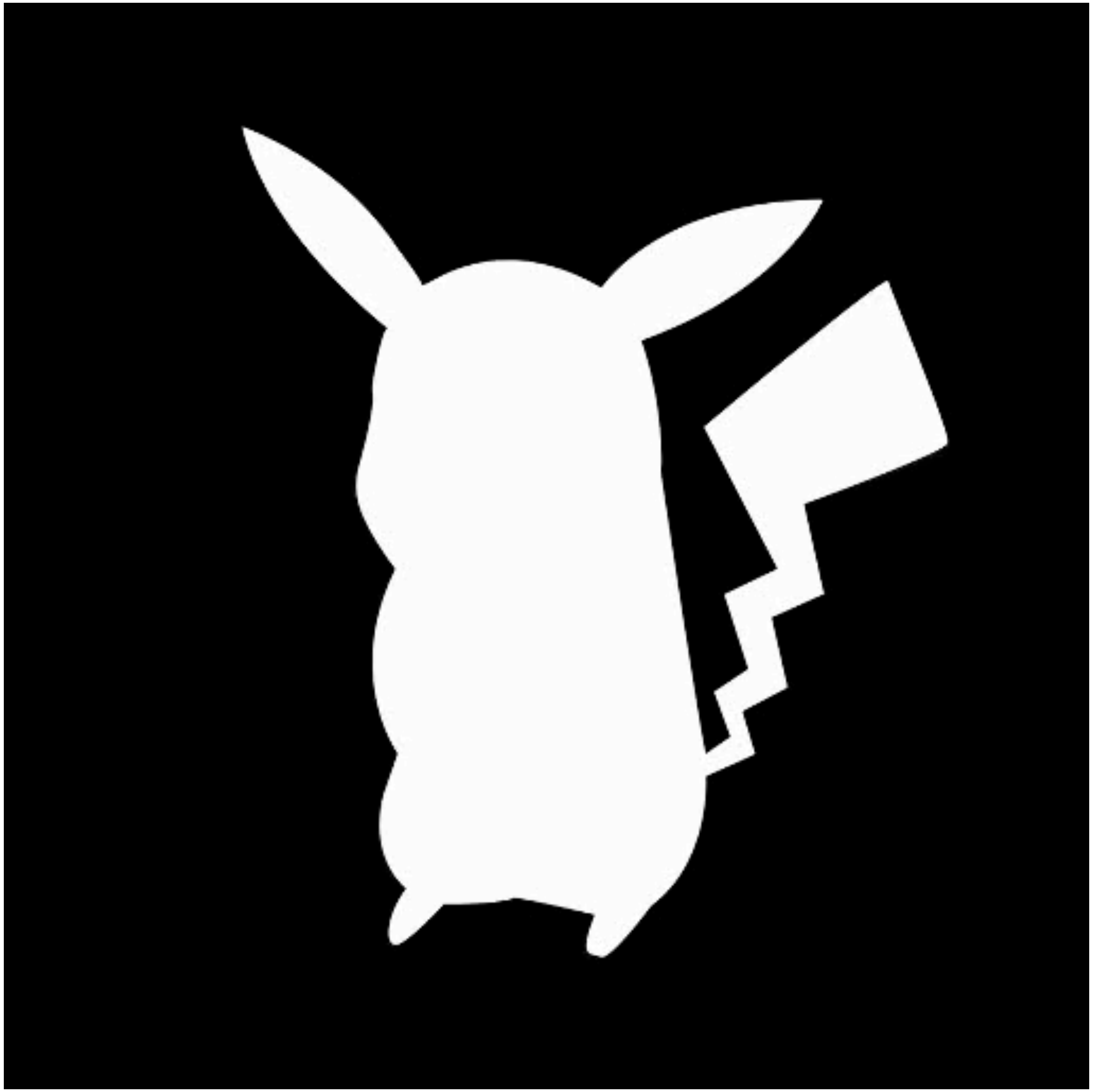
$$T(x) = \nabla\Psi(x).$$

This means

$$\det(\mathrm{Hess}\Psi(x))\rho_1(\nabla\Psi(x)) = \rho_0(x).$$

It is the Monge-Ampere equation, which is hard to solve directly.

Dynamical formulation



Benamou–Brenier

The distance has an important fluid dynamics formulation (Benamou-Brenier 2000). Then the square of L^2 Kantorovich distance satisfies

$$(W_2(\rho_0, \rho_1))^2 = \inf_{\rho, v} \int_0^1 \int \|v(t,x)\|^2 \rho(t,x) dx dt \ ,$$

where infimum runs over the continuity equation, such that

$$\partial_t \rho_t + \nabla \cdot (\rho v) = 0 \ , \quad \rho_0 = \rho_0 \ , \quad \rho_1 = \rho_1 \ .$$

Here the minimizer satisfies

$$v(t,x) = \nabla \Phi(t,x),$$

and

$$\frac{\partial}{\partial t} \Phi(t,x) + \frac{1}{2} \|\nabla \Phi(t,x)\|^2 = 0.$$

We shall focus on this formulation, and further propose an extension.

Unnormalized Optimal Transport

Define

$$UW_p(\mu_0, \mu_1)^p = \inf_{v,\mu,f} \int_0^1 \int_\Omega \|v(t,x)\|^p \mu(t,x) dx dt + \frac{1}{\alpha} \int_0^1 |f(t)|^p dt \cdot |\Omega|$$

such that the dynamical constraint: the unnormalized continuity equation holds

$$\partial_t \mu(t,x) + \nabla \cdot (\mu(t,x) v(t,x)) = f(t),$$

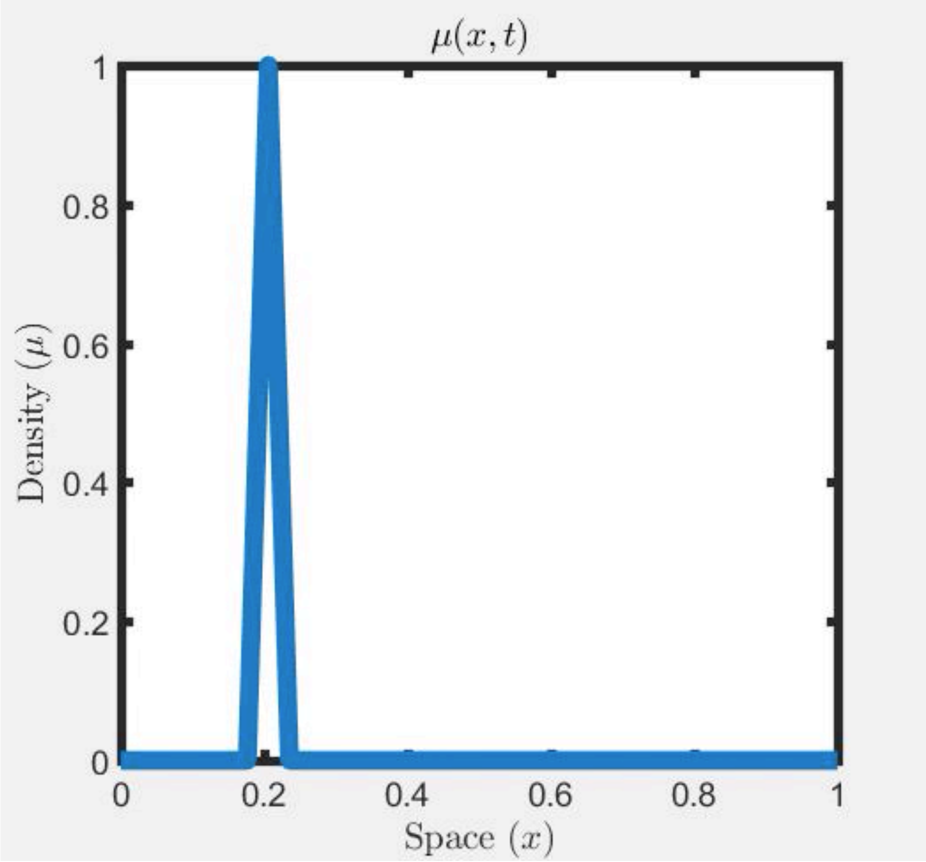
with

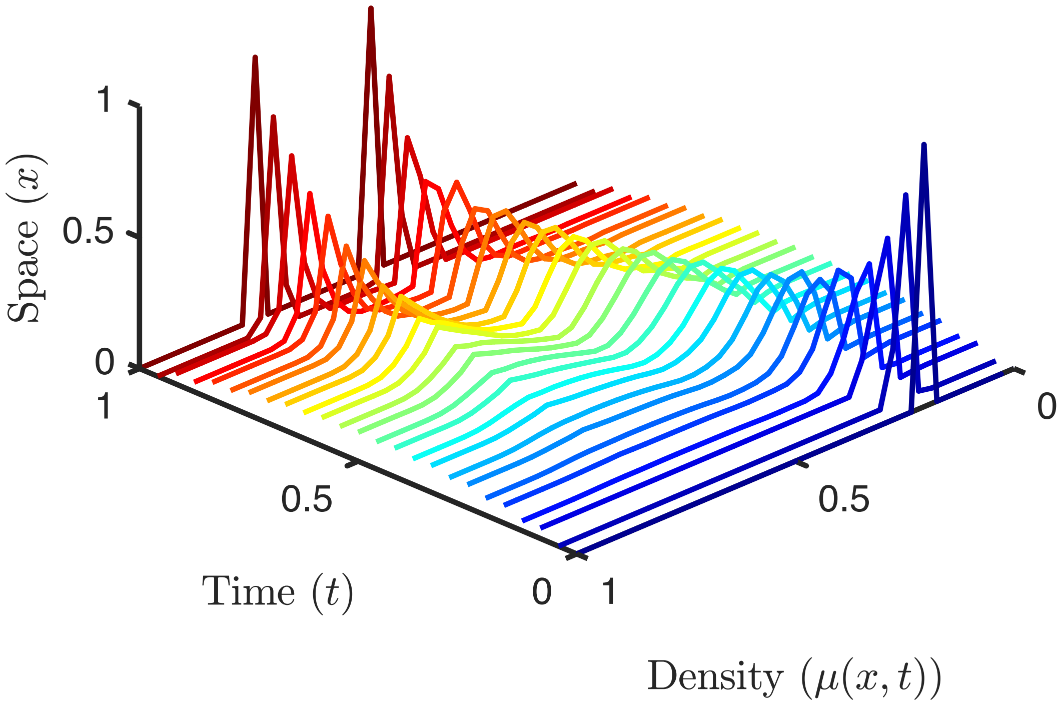
$$\mu(0,x) = \mu_0(x), \quad \mu(1,x) = \mu_1(x).$$

In this talk, we mainly consider $p = 1, 2$.

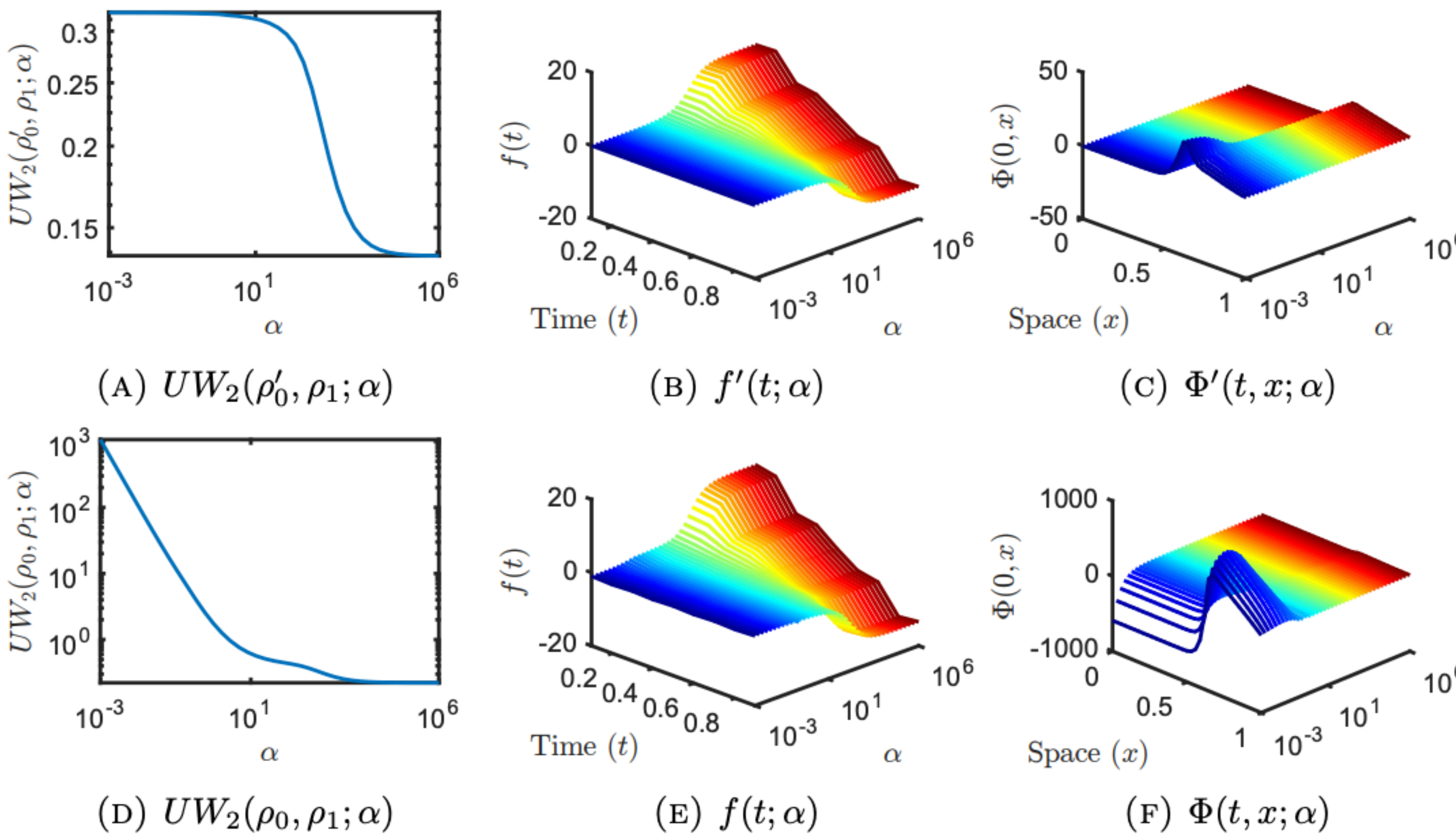
Snowing and melting

The source function $f(t)$ introduce the precisely co-dimensional one variation into the density space.





New metric



(A) $UW_2(\rho_0', \rho_1; \alpha)$ (B) $f'(t; \alpha)$ (C) $\Phi'(t, x; \alpha)$

(D) $UW_2(\rho_0, \rho_1; \alpha)$ (E) $f(t; \alpha)$ (F) $\Phi(t, x; \alpha)$

FIGURE 2. A plot of the asymptotic behavior of UW_2 in α with balanced and unbalanced inputs. Balanced: (A) $UW_2(\rho_0', \rho_1; \alpha)$, (B) $f'(t; \alpha)$, (C) $\Phi'(t, x; \alpha)$, and unbalanced: (D) $UW_2(\rho_0, \rho_1; \alpha)$, (E) $f(t; \alpha)$, (F) $\Phi(t, x; \alpha)$.

Unnormalized L^1 Wasserstein metric

Let $p = 1$:

$$\mathrm{UW}_1(\mu_0, \mu_1) = \inf_{v, f(t)} \Big\{ \int_0^1 \int_\Omega \|v\| \mu dx dt + \frac{1}{\alpha} \int_0^1 |f(t)| dt \cdot |\Omega| : \\ \partial_t \mu + \nabla \cdot (\mu v) = f(t) \Big\}.$$

Time independent solution

Denote

$$m(x) = \int_0^1 v(t,x)\mu(t,x)dt,$$

with the fact

$$\int_0^1 f(t)dt = c = \frac{1}{|\Omega|}\Big(\int_\Omega \mu_1(x)dx - \int_\Omega \mu_0(x)dx\Big).$$

then by Jensen's inequality and integrating the time variable t, we obtain

$$\begin{aligned}
&\mathrm{UW}_1(\mu_0,\mu_1)\\
=&\inf_m \Big\{ \int_\Omega \|m(x)\|dx + \frac{1}{\alpha}\Big|\int_\Omega \mu_0(x)dx - \int_\Omega \mu_1(x)dx\Big| :\\
&\quad \mu_1(x) - \mu_0(x) + \nabla\cdot m(x) = \frac{1}{|\Omega|}\Big(\int_\Omega \mu_1(x)dx - \int_\Omega \mu_0(x)dx\Big)\Big\}.
\end{aligned}$$

Closed form solution

In one space dimension on the interval $\Omega = [0,1]$, the L^1 unnormalized Wasserstein metric has the following explicit solution:

$$\begin{aligned}\mathrm{UW}_1(\mu_0,\mu_1) = &\int_\Omega \Big| \int_0^x \mu_1(y)dy - \int_0^x \mu_0(y)dy - x\int_\Omega (\mu_1(z)-\mu_0(z))dz \Big| dx \\ &+ \frac{1}{\alpha}\Big(\Big|\int_\Omega \mu_1(z)dz - \int_\Omega \mu_0(z)dz\Big|\Big).\end{aligned}$$

Algorithm

In high dimensional sample space, the L_1 unnormalized OT problem forms

$$\underset{m}{\text{minimize}} \quad \|m\|_{1,2} \quad \text{subject to} \quad \mathrm{div}(m) + \mu^1 - \mu^0 = c.$$

It is a particular example of compressed sensing. It can be solved easily by Primal-Dual algorithm (Chambolle and Pock).

Primal-Dual updates

Consider the Lagrangian of UOT:

$$\mathcal{L}(m, \Phi) = \int_0^1 \int_\Omega \|m\| + \Phi(\mathrm{div}(m) + \mu^1 - \mu^0) dx,$$

where $\Phi(t, x)$ is the Lagrange multiplier of the unnormalized continuity equation. The primal-dual update forms

$$\begin{cases} m^{k+1}(t, x) = \arg\inf_m \ \mathcal{L} + \dfrac{1}{2\tau_1} \displaystyle\int_0^1 \int_\Omega \|m(t, x) - m^k(t, x)\|^2 dx dt \\ \tilde{\Phi}^{k+1}(t, x) = \arg\sup_\Phi \ \mathcal{L} - \dfrac{1}{2\tau_2} \displaystyle\int_0^1 \int_\Omega \|\Phi(t, x) - \Phi^k(t, x)\|^2 dx dt \end{cases}$$

where m, Φ are taking the gradient descent, ascent directions respectively, with τ_1, τ_2 being the stepsizes.

Algorithm: 2 line codes

Primal-dual method

1. For $k = 1, 2, \cdots$ $\qquad$ Iterates until convergence
2. $\qquad m^{k+1} = \mathrm{shrink}(m^k + \mu \nabla \Phi^k, \mu)$;
3. $\qquad \Phi^{k+1} = \Phi^k + \tau\{\mathrm{div}(2m^{k+1} - m^k) + p^1 - p^0 + c\}$;
4. End

Here the shrink operator for the ground metric

$$\mathrm{shrink}(y, \alpha) := \frac{y}{\|y\|} \max\{\|y\| - \alpha, 0\} \ , \quad \text{where } y \in \mathbb{R}^d \ .$$

Examples

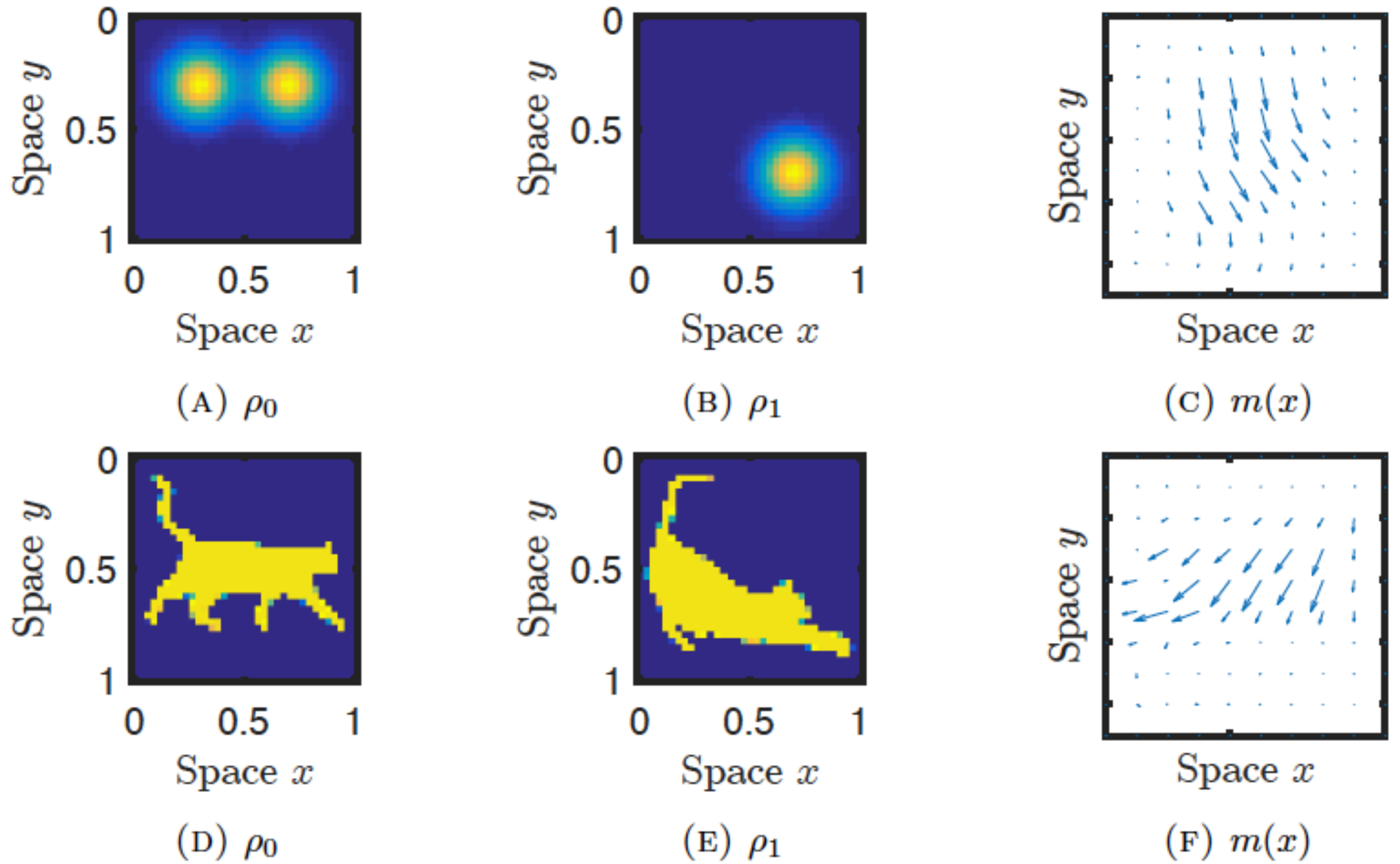


FIGURE 5. Plots of the ρ_0, ρ_1 and $m(x)$ for $UW_1(\rho_0, \rho_1)$ for the two gaussian movement (A) ρ_0, (B) ρ_1, (C) $m(x)$ and two (D) ρ_0, (E) ρ_1, (F) $m(x)$.

Unnormalized L^2 Wasserstein metric

Let $p = 2$:

$$\mathrm{UW}_2(\mu_0, \mu_1) = \inf_{v, f(t)} \Big\{ \int_0^1 \int_\Omega \|v\|^2 \mu dx dt + \frac{1}{\alpha} \int_0^1 f(t)^2 dt \cdot |\Omega| : \\ \partial_t \mu + \nabla \cdot (\mu v) = f(t) \Big\}.$$

Minimizer system

The minimizer $(v(t,x), \mu(t,x), f(t))$ for UOT problem satisfies

$$v(t,x) = \nabla\Phi(t,x), \quad f(t) = \alpha \frac{1}{|\Omega|} \int_\Omega \Phi(t,x)dx,$$

and

$$\begin{cases} \partial_t \mu(t,x) + \nabla \cdot (\mu(t,x)\nabla\Phi(t,x)) = \alpha \frac{1}{|\Omega|} \int_\Omega \Phi(t,x)dx \\ \partial_t \Phi(t,x) + \frac{1}{2}\|\nabla\Phi(t,x)\|^2 = 0 \\ \mu(0,x) = \mu_0(x), \quad \mu(1,x) = \mu_1(x). \end{cases}$$

Unnormalized Monge-Ampere equation

Denote

$$\Psi(x) = \frac{1}{2}\|x\|^2 + \Phi(0, x),$$

Following the Hopf-Lax formula, the minimizer of unnormalized OT satisfies

$$\begin{aligned}
&\mu(1, \nabla\Psi(x))\mathrm{Det}(\nabla^2\Psi(x)) - \mu(0, x) \\
=&\alpha\int_0^1 \mathrm{Det}\Big(t\nabla^2\Psi(x) + (1-t)\mathbb{I}\Big)\cdot \\
&\int_\Omega \Big(\Psi(y) - \frac{\|y\|^2}{2} + \frac{t\|\nabla\Psi(y) - y\|^2}{2}\Big)\mathrm{Det}\Big(t\nabla^2\Psi(y) + (1-t)\mathbb{I}\Big)dydt.
\end{aligned}$$

Unnormalized Kantorovich problem

$$\frac{1}{2}\mathrm{UW}_2(\mu_0,\mu_1)^2 = \sup_{\Phi} \left\{ \int_\Omega \Phi(1,x)\mu(1,x)dx - \int_\Omega \Phi(0,x)\mu(0,x)dx \right.$$
$$\left. - \frac{\alpha}{2}\int_0^1 \left(\int_\Omega \Phi(t,x)dx \right)^2 dt \right\}$$

where the supremum is taken among all $\Phi \colon [0,1] \to \Omega$ satisfying

$$\partial_t \Phi(t,x) + \frac{1}{2}\|\nabla \Phi(t,x)\|^2 \leq 0.$$

Algorithm

Denote $m(t,x) = \mu(t,x)v(t,x)$. Consider the Lagrangian of UOT:

$$\mathcal{L}(m,\mu,f,\Phi) = \int_0^1 \int_\Omega \frac{\|m(t,x)\|^2}{2\mu(t,x)} dtdx + \frac{1}{2\alpha}\int_0^1 f(t)^2 dt + \int_0^1 \int_\Omega \Phi(t,x)\Big(\partial_t \mu(t,x) + \nabla \cdot m(t,x) - f(t)\Big) dxdt,$$

where $\Phi(t,x)$ is the Lagrange multiplier of the unnormalized continuity equation. This formulation allows us to apply the primal dual algorithm for

$$\inf_{m,\mu} \sup_{f,\Phi} \mathcal{L}(m,\mu,f,\Phi).$$

Primal-Dual updates

$$
\begin{cases}
m^{k+1}(t,x) = \arg\inf\limits_{m} \ \mathcal{L} + \dfrac{1}{2\tau_1}\displaystyle\int_0^1\int_\Omega \|m(t,x) - m^k(t,x)\|^2 dx dt \\
\mu^{k+1}(t,x) = \arg\inf\limits_{\mu} \ \mathcal{L} + \dfrac{1}{2\tau_1}\displaystyle\int_0^1\int_\Omega \|\mu(t,x) - \mu^k(t,x)\|^2 dx dt \\
f^{k+1}(t) = \arg\inf\limits_{f} \ \mathcal{L} + \dfrac{1}{2\tau_1}\displaystyle\int_0^1 \|f(t) - f^k(t)\|^2 dt \\
\tilde{\Phi}^{k+1}(t,x) = \arg\sup\limits_{\Phi} \ \mathcal{L} - \dfrac{1}{2\tau_2}\displaystyle\int_0^1\int_\Omega \|\Phi(t,x) - \Phi^k(t,x)\|^2 dx dt \\
(\tilde{m}, \tilde{\mu}, \tilde{f}) = 2(m^{k+1}, \mu^{k+1}, f^{k+1}) - (m^k, \mu^k, f^k)
\end{cases}
$$

Algorithm

Algorithm: Primal-Dual method for Unnormalized OT

1. **For** $k = 1, 2, \cdots$ Iterate until convergence
2. $m^{k+1}(t,x) = \frac{\mu^k(t,x)}{\mu^k(t,x)+\tau_1}\Big(\tau_1 \nabla \Phi(t,x) + m^k(t,x)\Big)$;
3. $\mu^{k+1}(t,x) = \arg\inf_\mu \left(\frac{\|m^k\|^2}{2\mu} - \partial_t \Phi \cdot \mu + \frac{1}{2\tau_1}|\mu - \mu^k|^2\right)(t,x)$;
4. $f^{k+1}(t) = \frac{\alpha}{\alpha+\tau_1}\Big(\tau_1 \int_\Omega \Phi(t,x)dx + f^k(t)\Big)$;
5. $\Phi^{k+1}(t,x) = \Phi^k(t,x) + \tau_2\Big(\partial_t \tilde{\mu}^{k+1}(t,x) + \nabla \cdot \tilde{m}(t,x) - \tilde{f}(t)\Big)$;
6. $(\tilde{m}, \tilde{\mu}, \tilde{f}) = 2(m^{k+1}, \mu^{k+1}, f^{k+1}) - (m^k, \mu^k, f^k)$;
7. **end**

Example I

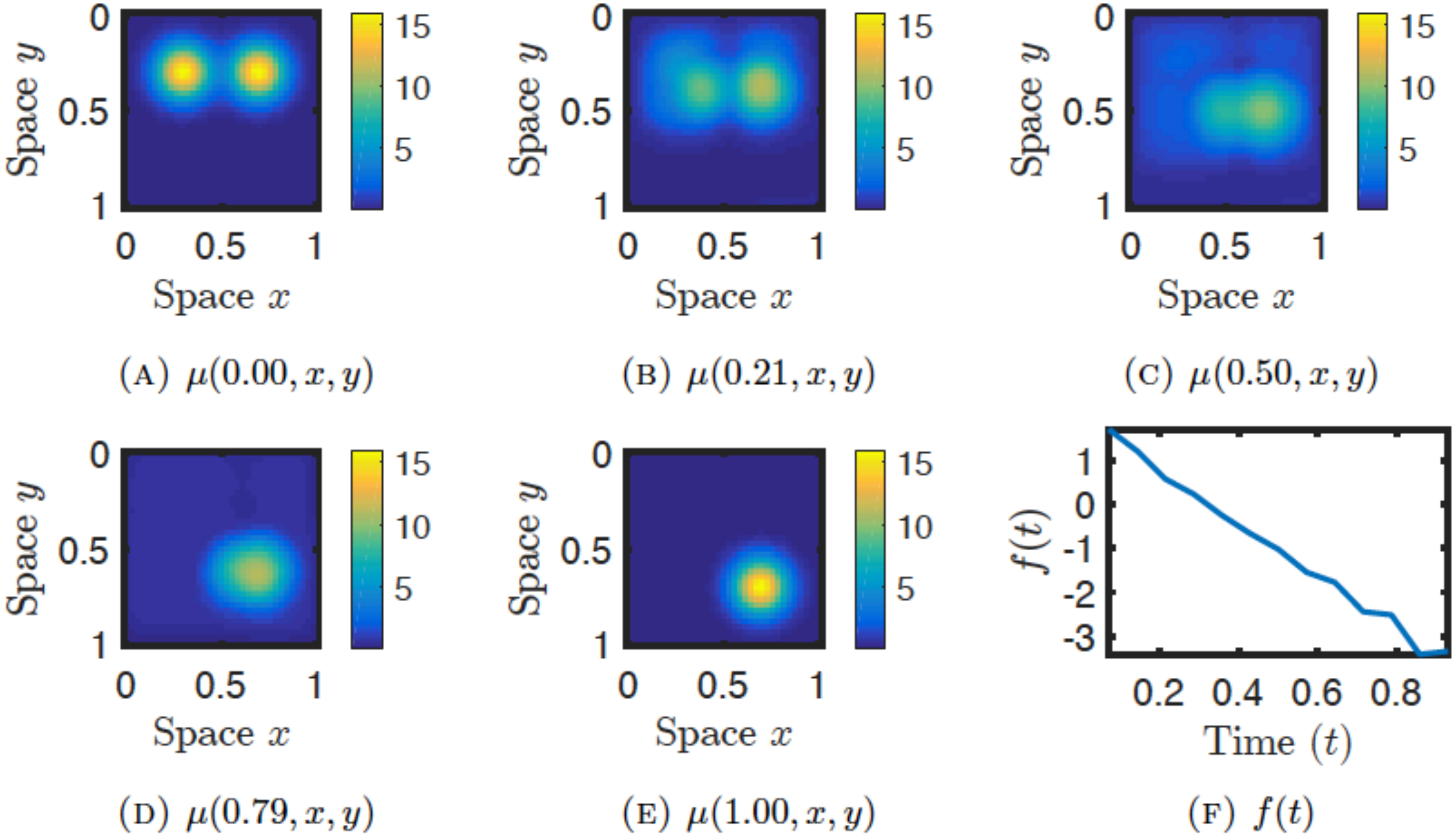


FIGURE 3. Plots of the $u(t, x, y)$ and $f(t)$ for $UW_2(\rho_0, \rho_1)$. (A) $\mu(0.00, x, y)$, (B) $\mu(0.21, x, y)$, (C) $\mu(0.50, x, y)$, (D) $\mu(0.79, x, y)$, (E) $\mu(1.00, x, y)$, (F) $f(t)$.

Example II

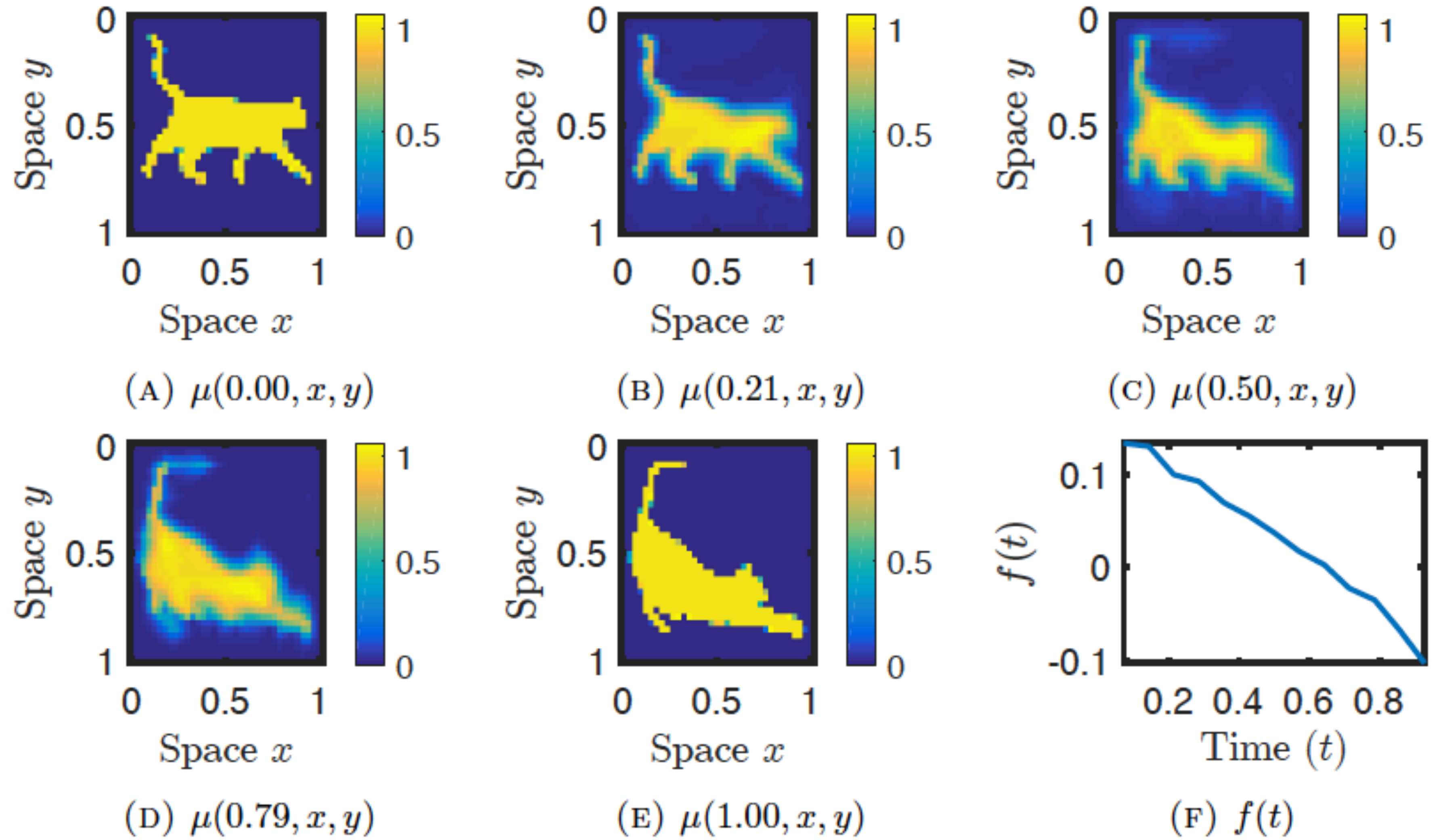
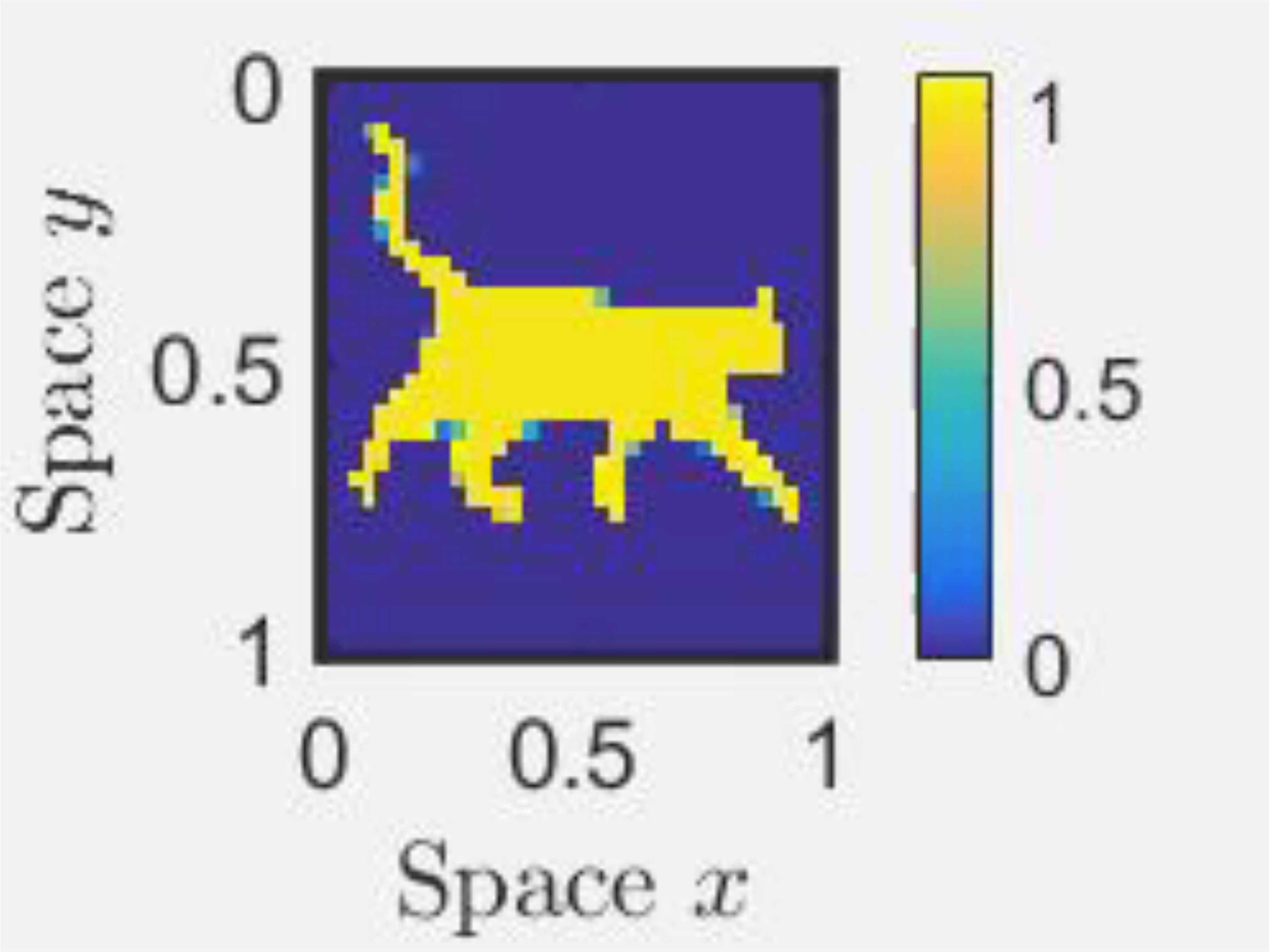


FIGURE 4. Plots of the $u(t, x, y)$ and $f(t)$ for $UW_2(\rho_0, \rho_1)$. (A) $\mu(0.00, x, y)$, (B) $\mu(0.21, x, y)$, (C) $\mu(0.50, x, y)$, (D) $\mu(0.79, x, y)$, (E) $\mu(1.00, x, y)$, (F) $f(t)$.

Example



Discussions

The unnormalized OT opens many interesting fields:

- Finding closed-form solutions of unnormalized OT;
- Modeling inverse problem via unnormalized OT;
- Geometric properties of unnormalized OT;
- Gradient flows via unnormalized OT;
- Mean field games and control problems in unnormalized density space.

Main references

W. Gangbo, W. Li, S. Osher and M. Puthawala.
Unnormalized Optimal Transport, 2019.

M. A. Puthawala, C. D. Hauck, and S. J. Osher.
Diagnosing Forward Operator Error Using Optimal Transport. 2018.

Y. Chow, W. Li, S. Osher and W. Yin.
Algorithm for Hamilton-Jacobi equations in density space via a generalized Hopf formula, 2018.

W. Li, P. Yin, and Stanley Osher.
Computations of optimal transport distance with Fisher information regularization, 2018.

W. Li, E. Ryu, S. Osher, W. Yin and W. Gangbo.
A parallel method for Earth mover's distance, 2017.