

Optimal Mass Transport and (matrix) density flows

Yongxin Chen (MSKCC & soon ISU)

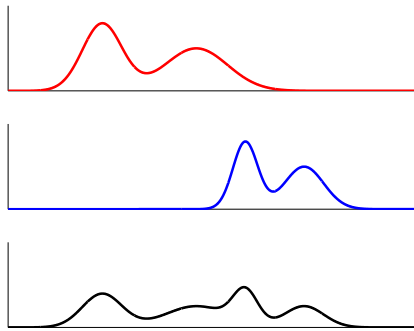
Tryphon Georgiou (Univ. of California, Irvine)

Allen Tannenbaum (Stony Brook)

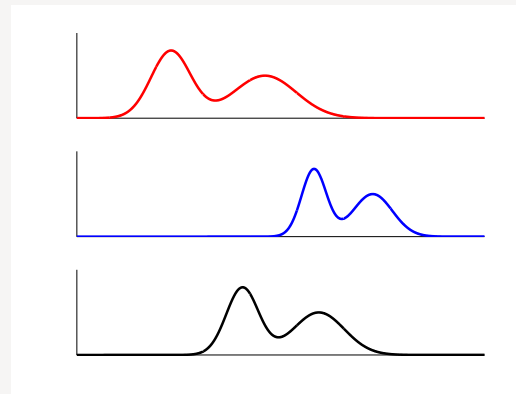
Optimal Transport meets
Probability, Statistics and Machine Learning
BIRS, Oaxaca

1 May - 5 May 2017

Motivation interpolation of densities



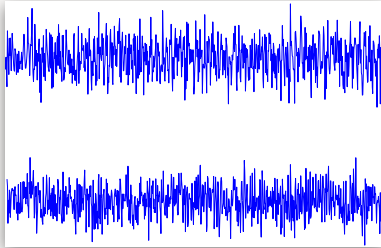
Arithmetic mean



Transportation mean

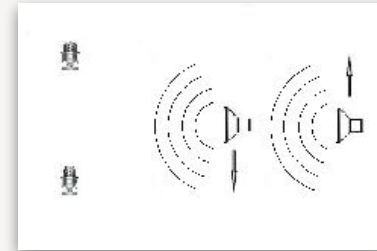
- push/pop?
- artifacts?
- etc

Time-series

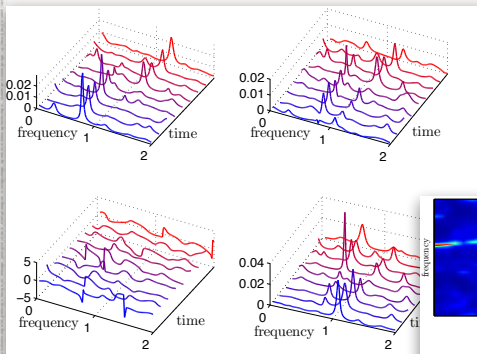


positioning
via
relative intensity
& doppler shift

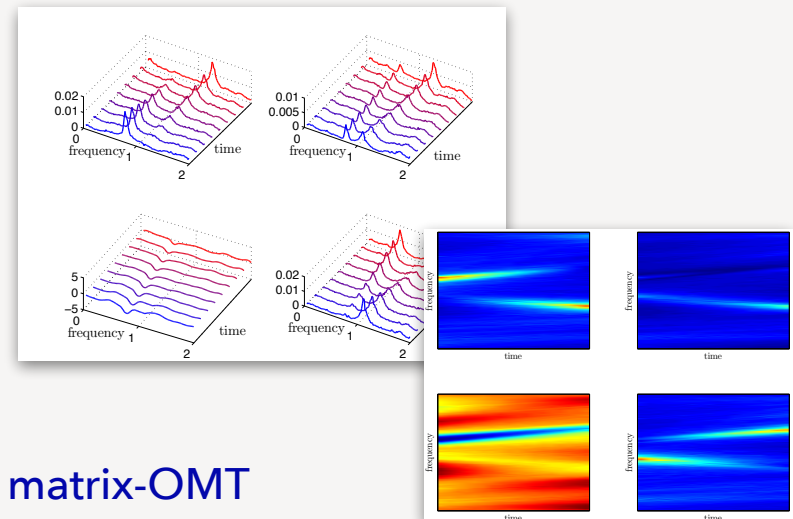
Physical arrangement



Matrix-valued power spectra

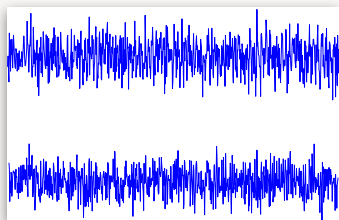


max-entropy
matrix-valued spectra

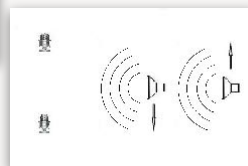


matrix-OMT
power spectral flow

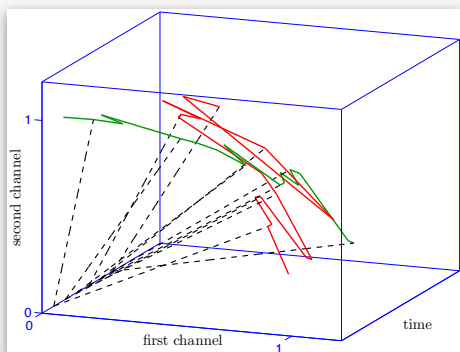
Time-series



Physical arrangement

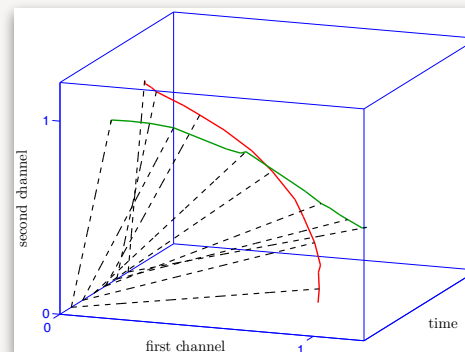


Time-flow of dominant eigen/singular vectors



eigenvectors of
max-entropy matrix-valued spectra

positioning
via
relative intensity
& doppler shift



eigenvectors of
matrix-OMT-like flow

Kantorovich-like formulation in product-space:

Ning, Georgiou, Tannenbaum, On matrix-valued Monge-Kantorovich OMT, 2015

OMT in quantum theory

Eric Carlen & Jan Maas

“An Analog of the 2-Wasserstein Metric..Fermionic Fokker-Planck.. Gradient Flow for the Entropy,” Comm. Math. Phys. 2014

arXiv:

Eric Carlen & Jan Maas

“Gradient flow and entropic inequalities...,” Sept 2016

Markus Mittnenzweig & Alexander Mielke

“An entropic gradient structure for Lindblad...,” Sept 2016.

Yongxin Chen, TTG & Allen Tannenbaum

“Matrix OMT: a Quantum Mechanical approach,” Oct 2016

Our goal

- extend the **Benamou-Brenier** framework to transport of
 - Hermitian matrices (Quantum density matrices)
 - matrix-valued distributions

I.e., formulate for matrices...

$$\inf \int \int_0^1 \rho(t, x) \|v(t, x)\|^2 dt dx$$

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho v) = 0,$$

$$\rho(0, \cdot) = \rho_0, \quad \rho(1, \cdot) = \rho_1$$

Quantum continuity equation

Starting point: Lindblad equation (in “diagonal form” $L_k = L_k^*$)

$$\begin{aligned}\dot{\rho} = & -[iH, \rho] \\ & + \sum_{k=1}^N (L_k \rho L_k - \frac{1}{2} \rho L_k L_k - \frac{1}{2} L_k L_k \rho),\end{aligned}$$

Notation:

\mathcal{H} and \mathcal{S} the set of $n \times n$ Hermitian and skew-Hermitian matrices

\mathcal{H}_+ and \mathcal{H}_{++} nonnegative and positive-definite matrices

$\mathcal{D}_+ := \{\rho \in \mathcal{H}_{++} \mid \text{tr}(\rho) = 1\}$ “density matrices”

$\mathcal{S}^N, \mathcal{H}^N$ block-column vectors with matrix-entries

Notation

$$\langle X, Y \rangle = \text{tr}(X^*Y), \quad X, Y \in \mathcal{H} \text{ (or } \mathcal{S})$$

$$\langle X, Y \rangle = \sum_{k=1}^N \text{tr}(X_k^* Y_k) \text{ for } X, Y \in \mathcal{H}^N \text{ (} \mathcal{S}^N)$$

For $X \in \mathcal{H}^N$ (or \mathcal{S}^N), $Y \in \mathcal{H}$ (or \mathcal{S}),

$$XY = \begin{bmatrix} X_1 \\ \vdots \\ X_N \end{bmatrix} Y := \begin{bmatrix} X_1 Y \\ \vdots \\ X_N Y \end{bmatrix},$$

and

$$YX = Y \begin{bmatrix} X_1 \\ \vdots \\ X_N \end{bmatrix} := \begin{bmatrix} Y X_1 \\ \vdots \\ Y X_N \end{bmatrix}.$$

Some calculus

Note for functions:

$$f(x) : g(x) \mapsto f(x)g(x)$$

$$\partial_x : g(x) \mapsto \partial_x g(x)$$

$$[\partial_x, f(x)] : g(x) \mapsto \partial_x f(x)g(x) - f(x)\partial_x g(x) = (\partial_x f(x))g(x)$$

For matrices:

$$\partial_{L_i} X = [L_i, X] = [L_i X - X L_i]$$

define the *gradient operator* for $L \in \mathcal{H}^N$

$$\nabla_L : \mathcal{H} \rightarrow \mathcal{S}^N, \quad X \mapsto \begin{bmatrix} L_1 X - X L_1 \\ \vdots \\ L_N X - X L_N \end{bmatrix}$$

Some calculus

∇_L is a derivation

$$\begin{aligned}\nabla_L(XY + YX) &= (\nabla_L X)Y + X(\nabla_L Y) \\ &\quad + (\nabla_L Y)X + Y(\nabla_L X), \quad \forall X, Y \in \mathcal{H}.\end{aligned}$$

dual is an analogue of the (negative) *divergence operator*:

$$\nabla_L^* : \mathcal{S}^N \rightarrow \mathcal{H}, \quad Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_N \end{bmatrix} \mapsto \sum_k^N L_k Y_k - Y_k L_k.$$

$$\langle \nabla_L X, Y \rangle = \langle X, \nabla_L^* Y \rangle$$

Lindblad term

Laplacian:

$$\begin{aligned}\Delta_L X &:= -\nabla_L^* \nabla_L X \\ &= \sum_{k=1}^N (2L_k X L_k - X L_k L_k - L_k L_k X), \quad X \in \mathcal{H},\end{aligned}$$

Lindblad's equation:

$$\dot{\rho} + [iH, \rho] = \sum_{k=1}^N (L_k \rho L_k - \frac{1}{2} \rho L_k L_k - \frac{1}{2} L_k L_k \rho),$$

becomes

$$\dot{\rho} + \nabla_{iH} \rho = \frac{1}{2} \Delta_L \rho.$$

Continuity equation

$$\dot{\rho} = \nabla_L^* M_\rho(v),$$

with $M_\rho(v)$ a “multiplication” between ρ and v
momentum field “ ρv ” = $M_\rho(v) \in \mathcal{S}^N$.

choices of non-commutative multiplication:

i) $\frac{1}{2}(\rho v + v \rho)$ (“anti-commutator”)

ii) $\int_0^1 \rho^s v \rho^{1-s} ds$ (Kubo-Mori)

iii) $\rho^{1/2} v \rho^{1/2}$

Case i) “anti-commutator”

Problem i):

$$W_{2,a}(\rho_0, \rho_1)^2 := \min_{\rho \in \mathcal{D}_+, v \in \mathcal{S}^N} \int_0^1 \text{tr}(\rho v^* v) dt,$$
$$\dot{\rho} = \frac{1}{2} \nabla_L^* (\rho v + v \rho),$$
$$\rho(0) = \rho_0, \quad \rho(1) = \rho_1,$$

Note: $v^* v = \sum_{k=1}^N v_k^* v_k$ and $v \in \mathcal{S}^N$.

Duality

$\lambda(\cdot) \in \mathcal{H}$ Lagrangian multiplier

$$\mathcal{L}(\rho, v, \lambda) = \int_0^1 \left\{ \frac{1}{2} \text{tr}(\rho v^* v) - \text{tr}(\lambda(\dot{\rho} - \frac{1}{2} \nabla_L^*(\rho v + v \rho))) \right\} dt$$

Point-wise minimization \Rightarrow

$$v_{opt}(t) = -\nabla_L \lambda(t).$$

Duality

If $\lambda(\cdot) \in \mathcal{H}$:

$$\dot{\lambda} = \frac{1}{2}(\nabla_L \lambda)^*(\nabla_L \lambda) = \frac{1}{2} \sum_{k=1}^N (\nabla_L \lambda)_k^* (\nabla_L \lambda)_k$$

and

$$\dot{\rho} = -\frac{1}{2} \nabla_L^* (\rho \nabla_L \lambda + \nabla_L \lambda \rho)$$

matches the marginals $\rho(0) = \rho_0, \rho(1) = \rho_1$,
then (ρ, v) with $v = -\nabla_L \lambda$ solves Problem i)

Riemannian structure

$$\delta_j \in \text{TangentSpace}_\rho = \{\delta \in \mathcal{H} \mid \text{tr}(\delta) = 0\}, \text{ for } j = 1, 2$$

“Poisson” equation: δ 's $\Leftrightarrow \lambda$'s

$$\delta_j = -\frac{1}{2} \nabla_L^* (\rho \nabla_L \lambda_j + \nabla_L \lambda_j \rho)$$

and

$$\langle \delta_1, \delta_2 \rangle_\rho = \frac{1}{2} \text{tr}(\rho \nabla \lambda_1^* \nabla \lambda_2 + \rho \nabla \lambda_2^* \nabla \lambda_1)$$

Note: given δ , then $-\nabla_L \lambda$ is the unique minimizer of $\text{tr}(\rho v^* v)$ over $v \in \mathcal{S}^N$ satisfying

$$\delta = \frac{1}{2} \nabla_L^* (\rho v + v \rho).$$

Riemannian metric

$W_{2,a}(\cdot, \cdot)$ is a metric on \mathcal{D}_+

$$W_{2,a}(\rho_0, \rho_1) = \min_{\rho} \int_0^1 \sqrt{\langle \dot{\rho}(t), \dot{\rho}(t) \rangle_{\rho(t)}} dt,$$

over piecewise smooth path on \mathcal{D}_+

Computation – convex problem

momentum field $u = \rho v$, i.e. $u_i = \rho v_i$

$$\mathrm{tr}(\rho v^* v) = \sum_{k=1}^N \mathrm{tr}(\rho v_k^* v_k) = \mathrm{tr}(u^* \rho^{-1} u),$$

define $u_* := [u_1, \dots, u_N]^*$, then

$$\begin{aligned} W_{2,a}(\rho_0, \rho_1)^2 &= \min_{\rho, u} \int_0^1 \mathrm{tr}(u^* \rho^{-1} u) dt, \\ \dot{\rho} &= \frac{1}{2} \nabla_L^* (u - u_*), \\ \rho(0) &= \rho_0, \quad \rho(1) = \rho_1 \end{aligned}$$

Note: optimal u automatically satisfies $u = \rho v$ for some $v \in \mathcal{S}^N$
no need for a constraint

Matrix transport with added spatial component

$$\mathcal{D} = \{\rho(\cdot) \mid \rho(x) \in \mathcal{H}_+ \text{ such that } \int_{\mathbb{R}^m} \text{tr}(\rho(x)) dx = 1\}.$$

Continuity equation: $w \in \mathcal{H}$ along space dimension

$$\frac{\partial \rho}{\partial t} + \frac{1}{2} \nabla_x \cdot (\rho w + w \rho) - \frac{1}{2} \nabla_L^* (\rho v + v \rho) = 0.$$

Metric:

$$W_{2,a}(\rho_0, \rho_1)^2 := \min \int_0^1 \int_{\mathbb{R}^m} \{\text{tr}(\rho w^* w) + \gamma \text{tr}(\rho v^* v)\} dx dt,$$
$$\rho \in \mathcal{D}_+, w \in \mathcal{H}^m, v \in \mathcal{S}^N$$
$$\frac{\partial \rho}{\partial t} + \frac{1}{2} \nabla_x \cdot (\rho w + w \rho) - \frac{1}{2} \nabla_L^* (\rho v + v \rho) = 0,$$
$$\rho(0, \cdot) = \rho_0, \quad \rho(1, \cdot) = \rho_1$$

Metric, computation, etc.

Same as before:

duality..

$$w_{\text{opt}}(t, x) = -\nabla_x \lambda(t, x), \quad v_{\text{opt}} = -\frac{1}{\gamma} \nabla_L \lambda(t, x)$$

$$\delta_j \xrightarrow{\text{Poisson}} \lambda_j \text{'s}$$

and then $\langle \delta_1, \delta_2 \rangle = \int$ “symmetrized kinetic energy” dx

metric computed via **convex optimization**, with $q = \rho w$, $u = \rho v$:

$$\begin{aligned} & \min_{\rho, q, u} \int_0^1 \int_{\mathbb{R}^m} \{ \text{tr}(q^* \rho^{-1} q) + \gamma \text{tr}(u^* \rho^{-1} u) \} dx dt \\ & \frac{\partial \rho}{\partial t} + \frac{1}{2} \nabla_x \cdot (q + q^*) - \frac{1}{2} \nabla_L^* (u - u^*) = 0, \\ & \rho(0, \cdot) = \rho_0, \quad \rho(1, \cdot) = \rho_1 \end{aligned}$$

Gradient flow of Entropy

$$S(\rho) = -\text{tr}(\rho \log \rho).$$

then

$$\begin{aligned}\frac{dS(\rho(t))}{dt} &= \dots \\ &= -\frac{1}{2} \text{tr}(\rho v^* \nabla_L \log \rho + \rho (\nabla_L \log \rho)^* v),\end{aligned}$$

\Rightarrow steepest ascent

$$v = -\nabla_L \log \rho$$

\Rightarrow **gradient flow**

$$\dot{\rho} = -\frac{1}{2} \nabla_L^* \{ \rho, \nabla_L \log \rho \}$$

Note: this is nonlinear, different from Lindblad

Note: similar with space component..

Case ii) “logarithmic”

Problem ii):

$$W_{2,b}(\rho_0, \rho_1)^2 := \min_{\rho \in \mathcal{D}_+, v \in \mathcal{S}^N} \int_0^1 \int_0^1 \text{tr}(v^* \rho^s v \rho^{1-s}) ds dt$$
$$\dot{\rho} = \nabla_L^* \int_0^1 \rho^s v \rho^{1-s} ds,$$
$$\rho(0) = \rho_0, \quad \rho(1) = \rho_1.$$

Note: computations?

duality

If $\lambda(\cdot) \in \mathcal{H}$, ρ satisfy

$$\dot{\lambda} = \int_0^1 \int_0^1 \int_0^\alpha \left\{ \frac{\rho^{\alpha-\beta}}{(1-s)I + s\rho} (\nabla_L \lambda)^* \rho^{1-\alpha} \nabla_L \lambda \frac{\rho^\beta}{(1-s)I + s\rho} \right\} d\beta d\alpha ds$$

$$\dot{\rho} + \nabla_L^* \int_0^1 \rho^s \nabla_L(\lambda) \rho^{1-s} ds = 0,$$

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

then $(\rho, -\nabla_L(\lambda))$ is optimal.

Gradient flow of Entropy

$$\begin{aligned}\frac{dS(\rho(t))}{dt} &= \dots \\ &= -\text{tr}((\nabla_L \log \rho)^* \int_0^1 \rho^s v \rho^{1-s} ds),\end{aligned}$$

\Rightarrow greatest ascent direction $v = -\nabla_L \log \rho$.

non-commutative analog of: $\partial_x \rho = \rho \partial_x (\log \rho)$:

$$\nabla_L \rho = \int_0^1 \rho^s (\nabla_L \log \rho) \rho^{1-s} ds$$

Gradient flow:

$$\dot{\rho} = -\nabla_L^* \int_0^1 \rho^s (\nabla_L \log \rho) \rho^{1-s} ds = -\nabla_L^* \nabla_L \rho = \Delta_L \rho,$$

Linear heat equation (now Lindblad) just as in the scalar case!

Recap

With $M_\rho(v) = \rho v + v \rho$:

metric computable via convex optimization

gradient flow of entropy: nonlinear

With $M_\rho(v) = \int_0^1 \rho^s v \rho^{1-s} ds$:

metric computability questionable

gradient flow of entropy: linear, Lindblad

Strong duality & conservation of Hamiltonian

Y. Chen, Wilfrid Gangbo, TTG & A. Tannenbaum

“On the Matrix Monge-Kantorovich Problem, arXiv 2017

Strong duality: define $F(\rho, m) := \frac{1}{2} \langle m, m \rho^{-1} \rangle$

Let $\rho_0, \rho_1 \in \mathcal{D}_+$. Then

$$\begin{aligned} & \min_{(\rho, m) \in \mathcal{A}} \left\{ \int_0^1 F(\rho, m) dt \mid \rho(0) = \rho_0, \rho(1) = \rho_1, \text{ and } \dot{\rho} = \frac{1}{2} \nabla_L^* (m - m_*) \right\} \\ & = \sup_{\lambda \in \mathcal{B}} \left\{ \langle \lambda(1); \rho_1 \rangle - \langle \lambda(0); \rho_0 \rangle \mid \dot{\lambda} + \frac{1}{2} (\nabla_L \lambda)^* (\nabla_L \lambda) \leq 0 \text{ a.e. on } (0, 1) \right\}. \end{aligned}$$

$$\mathcal{A} := \{ \rho \in L^2(0, 1; \mathcal{H}) \mid \text{tr}(\rho) \equiv 1 \} \times L^2(0, 1; \mathbb{C}^{nN \times n})$$

$$\mathcal{B} := W^{1,2}(0, 1; \mathcal{H})$$

Conservation of the Hamiltonian:

Let $\rho_0, \rho_1 \in \mathcal{D}_+$ and (ρ, m) a minimizer as before. Then:

(i)
$$F(\rho(t), m(t)) \equiv F(\rho(0), m(0)).$$

(ii) If $0 \leq s \leq t \leq 1$ then

$$W_2(\rho(s), \rho(t)) = (t - s) \sqrt{2F(\rho(t), m(t))} = (t - s) W_2(\rho_0, \rho_1).$$

(iii) If we further assume that $\lambda \in W^{1,1}(0, 1; \mathcal{H})$ is a maximizer of the dual, then

$$\langle \lambda(t); \rho(t) \rangle = \langle \lambda(0); \rho_0 \rangle + \frac{W_2(\rho_0, \rho(t))^2}{2t}, \quad t \in (0, 1].$$

unbalanced transport

J.-D. Benamou and Y. Brenier

“A computational fluid mechanics solution ..” L^2 and Wasserstein

L. Chizat, B. Schmitzer, G. Peyré, and F.-X. Vialard

“An interpolating ... optimal transport and Fisher-Rao”

S. Kondratyev, L. Monsaingeon, and D. Vorotnikov

“A new optimal transport distance...,” OMT and Fisher-Rao

M. Liero, A. Mielke, Giuseppe Savaré

“Optimal entropy-transport problems and a new Hellinger-Kantorovich distance..”

unbalanced: $\text{trace } \rho_0 \neq \text{trace } \rho_1$

arXiv: Y. Chen, TTG & A. Tannenbaum

“Interpolation of Matrix-Valued Measures: The Unbalanced Case

Interpolation between Wasserstein and Bures:

$$W_{2,FR}(\rho_0, \rho_1)^2 := \inf_{\rho \in \mathcal{H}_{++}, v \in \mathcal{S}^N, r \in \mathcal{H}} \int_0^1 \{ \text{tr}(\rho v^* v) + \alpha \text{tr}(\rho r^2) \} dt$$
$$\dot{\rho} = \frac{1}{2} \nabla_L^*(\rho v + v \rho) + \frac{1}{2}(\rho r + r \rho),$$
$$\rho(0) = \rho_0, \quad \rho(1) = \rho_1.$$

Interpolation between Wasserstein and Frobenius:

$$W_{2,F}(\rho_0, \rho_1)^2 := \inf_{\rho \in \mathcal{H}_{++}, v \in \mathcal{S}^N, s \in \mathcal{H}} \int_0^1 \{ \text{tr}(\rho v^* v) + \alpha \text{tr}(s^2) \} dt$$
$$\dot{\rho} = \frac{1}{2} \nabla_L^*(\rho v + v \rho) + s,$$
$$\rho(0) = \rho_0, \quad \rho(1) = \rho_1.$$

- can be turned into **convex problems** as usual...

transport of vector-valued distributions

$$\rho = [\rho_1, \rho_2, \dots, \rho_\ell]^T, \text{ on } \mathbb{R}_+^N$$

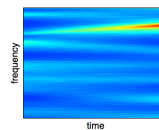
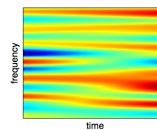
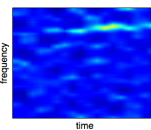
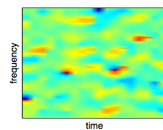
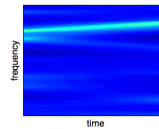
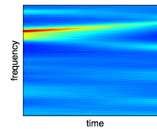
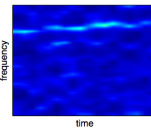
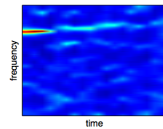
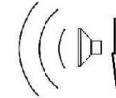
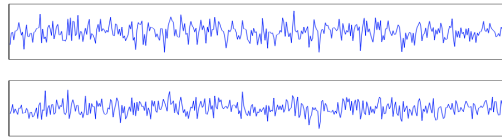
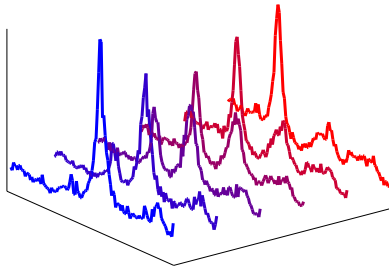
$$\sum_{i=1}^{\ell} \int_{\mathbb{R}^N} \rho_i(x) dx = 1,$$

continuity equation:

$$\frac{\partial \rho_i}{\partial t} + \nabla_x \cdot \underbrace{(\rho_i v_i)}_{u_i} - \sum_{j \neq i} \underbrace{(\rho_j w_{ji} - \rho_i w_{ij})}_{p_{ji}} = 0, \quad \forall i = 1, \dots, \ell.$$

$$W_2(\mu, \nu)^2 := \inf_{\rho, v, w} \int_0^1 \int_{\mathbb{R}^N} \left\{ \sum_{i=1}^{\ell} \rho_i(t, x) \|v_i(t, x)\|^2 + \gamma \sum_{i, j=1}^{\ell} \rho_i w_{ij}^2(t, x) \right\} dx dt$$

- flows and metrics
for matrix and vector-valued distributions
for problems in signal analysis



thank you for your attention