# Optimal transport for diffeomorphic registration

We define a fidelity term based on Optimal Transport to compare unlabeled shape data, and couple it we a registration algorithm.

Jean Feydy[1,2] Benjamin Charlier[3,5] François-Xavier Vialard[4,6] Gabriel Peyré[1,5]

[1]DMA – École Normale Supérieure, Paris, France [2]CMLA – ENS Cachan, Cachan, France
[3]Institut Montpelliérain Alexander Grothendieck, Univ. Montpellier, Montpellier, France
[4]Univ. Paris-Dauphine - PSL Research, Paris, France [5]CNRS, Paris, France [6]INRIA Mokaplan, Paris, France

## The Measure+Kernel Paradigm

We focus on the registration of a shape $A$ to a shape $B$ through a (rigid, diffeomorphic, etc.) transformation $\varphi$:

$$A \;\longrightarrow\; \varphi(A) \;\simeq\; B.$$

In a variational setting, one chooses a transformation $\varphi$ minimizing an energy

$$E(\varphi) \;=\; \underbrace{\text{Reg}(\varphi)}_{\text{Regularization}} \;+\; \underbrace{d(\varphi(A) \to B)}_{\text{fidelity term}}.$$

Registration toolboxes thus require fidelity routines $d$ between unlabeled shapes $\varphi(A)$ and $B$. Conveniently, one represents those as measures:

$$\varphi(A) \leftrightarrow \mu = \sum_{i=1}^{I} p_i \delta_{x_i} \quad \text{and} \quad B \leftrightarrow \nu = \sum_{j=1}^{J} q_j \delta_{y_j}.$$

- If $\varphi(A)$ is a segmented surface, each weighted dirac $p_i\delta_{x_i}$ stands for a triangle.
- If $\varphi(A)$ is a segmented density image, each weighted dirac $p_i\delta_{x_i}$ stands for a voxel.

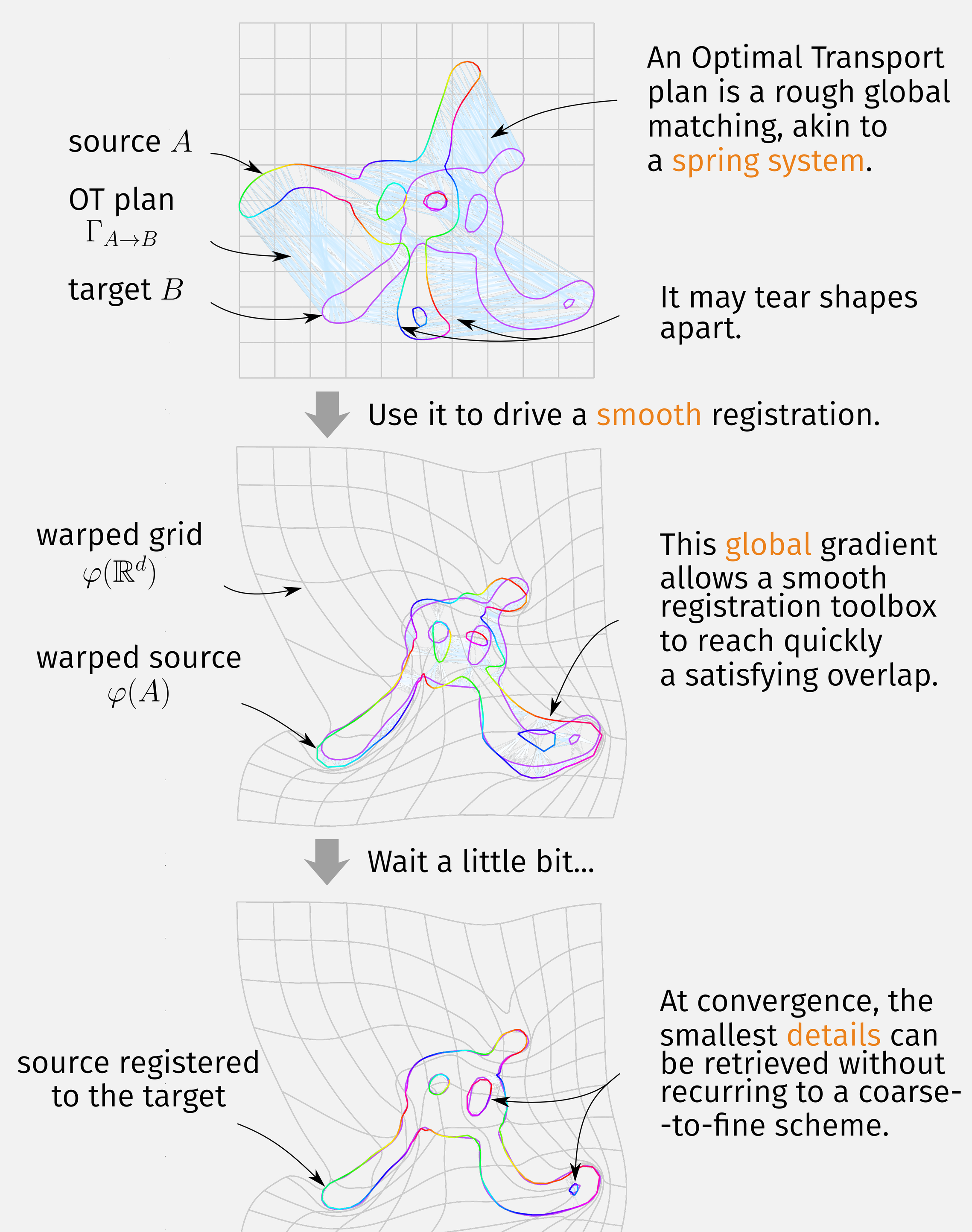
Then, one typically chooses a blurring function $G_\sigma$ associated to a kernel $k = G_\sigma \star G_\sigma$ and use

$$d(\mu \to \nu) \;=\; \| G_\sigma \star \mu - G_\sigma \star \nu \|^2_{L^2} = \langle \mu{-}\nu \,|\, k{\star}(\mu{-}\nu) \rangle.$$

This simple fidelity can be computed at the cost of a single convolution through the data $(\mu - \nu)$.

**Fig. 1 :** Smoothed data $G_\sigma \star (\mu - \nu)$ for two different scales $\sigma$. (a) Fine kernels are not suited to large deformations, whereas (b) heavy-tailed kernels can be hard to tune.

## Our Contribution




The proposed data attachment term is:

- Global, unlike kernel methods.
- Principled, as it relies on a blooming mathematical field.
- Differentiable, pluggable in any registration toolbox.
- Versatile, as it covers all scales and can be adapted to any feature space.
- Affordable, at a cost of 100-1000 gaussian convolutions per transport plan.

## The Math Behind It

In the simplest setting, we assume that

$$\mu = \sum_{i=1}^{I} p_i \delta_{x_i} \qquad \text{and} \qquad \nu = \sum_{j=1}^{J} q_j \delta_{y_j}$$

have the same total mass. Then, we define the Wasserstein fidelity term through a minimization on $I$-by-$J$ matrices – called transport plans $\Gamma$:

$$\underbrace{W_\varepsilon(\mu,\nu)}_{= d(\mu\to\nu)} \;=\; \min_{\Gamma} \underbrace{\sum_{i,j} \gamma_{i,j} \cdot |x_i - y_j|^2}_{\text{transport cost}} + \underbrace{\varepsilon \sum_{i,j} \gamma_{i,j} \log \gamma_{i,j}}_{\text{entropic regularization}}$$

under the constraint that $\Gamma = (\gamma_{i,j})$ satisfies

$$\forall i, j, \;\; \gamma_{i,j} \geqslant 0, \;\; \sum_j \gamma_{i,j} \;=\; p_i, \;\; \sum_i \gamma_{i,j} \;=\; q_j. \quad (1)$$

Optimality conditions show that the OT plan can be written as a product

$$\gamma_{i,j} \;=\; \Gamma(x_i \to y_j) \;=\; a(x_i)\, k(x_i, y_j)\, b(y_j), \quad \text{where:}$$

- The kernel function $k$ is given by
$$k(x_i, y_j) \;=\; k(x_i - y_j) \;=\; e^{-|x_i - y_j|^2/\varepsilon}.$$
- $a$ and $b$ are nonnegative functions supported respectively by $\{x_i\}$ and $\{y_j\}$.

The Sinkhorn theorem then asserts that $a$ and $b$ are uniquely determined by eq. (1), which now reads

$$a \;=\; \frac{p}{k \star b}, \qquad b \;=\; \frac{q}{k \star a}.$$

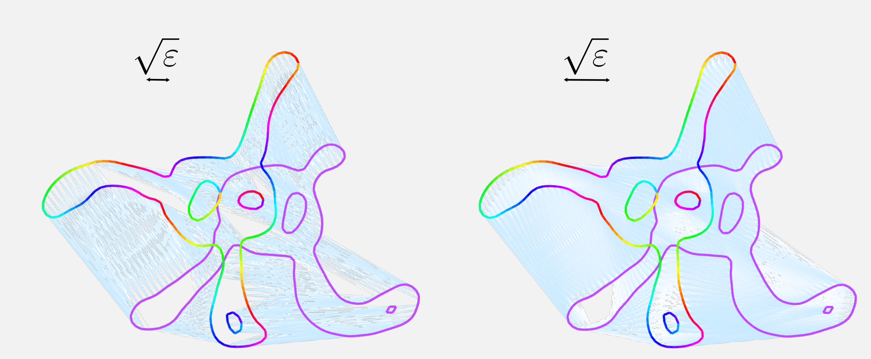



**Fig. 2 :** Two OT plans computed with different regularization scales $\sqrt{\varepsilon}$. Increasing this parameter results in a lower computational cost.

## The Algorithm

**Sinkhorn Iterative Algorithm**

**Parameter :** $k : x \mapsto e^{-|x|^2/\varepsilon}$
**Input :** source $\mu = \sum_i p_i \delta_{x_i}$
target $\nu = \sum_j q_j \delta_{y_j}$
**Output :** fidelity $W_\varepsilon(\mu,\nu)$

$a \leftarrow ones(size(p))$
$b \leftarrow ones(size(q))$
**while** updates $>$ tol **do**
  $a \leftarrow p \;/\; (k \star b)$
  $b \leftarrow q \;/\; (k \star a)$
**return** $\varepsilon \cdot \big( \langle p, \log(a) + 1/2 \rangle + \langle q, \log(b) + 1/2 \rangle \big)$

If the data lies on a grid, $k \star \cdot$ is a separable gaussian convolution.

If the data is sparse, $k \star \cdot$ is the product with the kernel matrix

$$(K_{ij}) \;=\; k(x_i, y_j)$$

and its transpose.

## In Practice

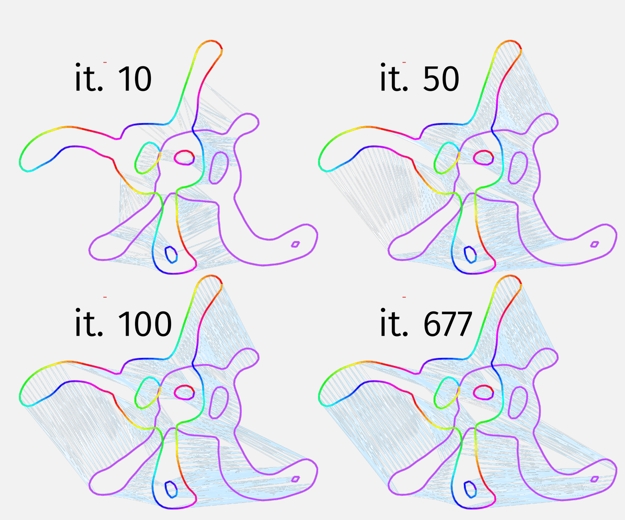



**Fig. 3 :** Sinkhorn iterations propagate along both shapes the information encoded within $(K_{ij})$.

The practical convergence rate of the Sinkhorn algorithm is not well understood yet, but computing an optimal transport plan typically requires ~1000 convolutions, depending on $\varepsilon$. Our Matlab and Python implementations are freely available: `github.com/jeanfeydy/lddmm-ot`

## Bonus Features

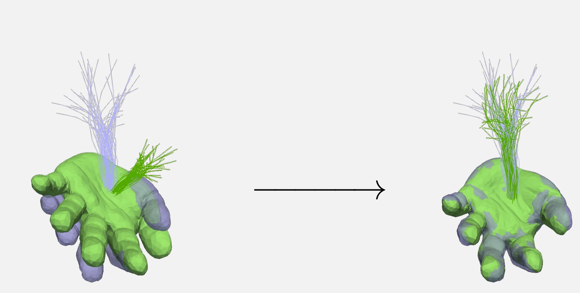


**Fig. 4 :** Use OT plans to registrate exotic data types.

- Use Unbalanced Transport, relaxing the constraints of eq. (1) with a soft penalty term.
- Generalize the algorithm to Features Spaces such as the “position + orientation” space.
- Compute seamlessly the derivatives of the fidelity.
- Implement the algorithm in the log-domain with Nesterov acceleration for increased numerical stability and speed.

## Take-Home Points

- The Sinkhorn algorithm, an iterative globalization trick, provides small kernels with long-distance vision.
- Computed at the cost of a few hundred convolutions, Optimal Transport plans can be used as spring systems driving a diffeomorphic registration routine.
- The resulting framework is more robust than a kernel-based one, as no target data is “out-of-sight”.
- This new scheme will find its use at the coarsest scales, where its properties are worth the computational overhead.

## References

1. Al-Rfou, R., Alain, G., Almahairi, A., Angermüller, C., Bahdanau, D., Ballas, N., ..., Zhang, Y.: Theano: A python framework for fast computation of mathematical expressions. CoRR abs/1605.02688 (2016)
2. Avants, B., Epstein, C., Grossman, M., Gee, J.: Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. Medical Image Analysis 12(1), 26 – 41 (2008)
3. Charon, N., Trouvé, A.: The varifold representation of nonoriented shapes for dif- feomorphic registration. SIAM J. Imaging Sciences 6(4), 2547–2580 (2013)
4. Chizat, L., Peyré, G., Schmitzer, B., Vialard, F.X.: Scaling algorithms for unbal- anced transport problems. Preprint 1607.05816, Arxiv (2016)
5. Chizat, L., Schmitzer, B., Peyré, G., Vialard, F.X.: An interpolating distance be- tween optimal transport and Fisher–Rao. to appear in Foundations of Computa- tional Mathematics (2016)
6. Cuturi, M.: Sinkhorn distances: Lightspeed computation of optimal transportation. In: Proc. NIPS, vol. 26, pp. 2292–2300 (2013)
7. Cuturi, M., Doucet, A.: Fast computation of Wasserstein barycenters. In: Proceed- ings of the 31st International Conference on Machine Learning (ICML), JMLR WCP. vol. 32 (2014)
8. Gee, J.C., Gee, J.C., Reivich, M., Reivich, M., Bajcsy, R., Bajcsy, R.: Elastically deforming a three-dimensional atlas to match anatomical brain images. J. Comput. Assist. Tomogr 17, 225–236 (1993)
9. Gori, P., Colliot, O., Marrakchi-Kacem, L., Worbe, Y., de Vico Fallani, F., Chavez, M., Lecomte, S., Poupon, C., Hartmann, A., Ayache, N., Durrleman, S.: A proto- type representation to approximate white matter bundles with weighted currents. Proceedings of Medical Image Computing and Computer Aided Intervention (MIC- CAI’14) Part III LNCS 8675, 289–296 (2014)
10. Liero, M., Mielke, A., Savaré, G.: Optimal entropy-transport problems and a new Hellinger–Kantorovich distance between positive measures. ArXiv e-prints (2015)
11. Montavon, G., Müller, K.R., Cuturi, M.: Wasserstein training of restricted Boltz- mann machines. In: Adv. in Neural Information Processing Systems (2016)
12. Santambrogio, F.: Optimal transport for applied mathematicians. Progress in Non- linear Differential Equations and their applications 87 (2015)
13. Sejdinovic, D., Sriperumbudur, B., Gretton, A., Fukumizu, K.: Equivalence of distance-based and rkhs-based statistics in hypothesis testing. Ann. Statist. 41(5), 2263–2291 (10 2013)
14. Sotiras, A., Davatzikos, C., Paragios, N.: Deformable medical image registration: A survey. IEEE Transactions on Medical Imaging 32(7), 1153–1190 (July 2013)
15. Vaillant, M., Glaunès, J.: Surface matching via currents. In: Biennial International Conference on Information Processing in Medical Imaging. pp. 381–392. Springer (2005)

## Contact Information

Jean Feydy,
PhD student with Prof. Alain Trouvé,
DMA, ENS Paris + CMLA, ENS Cachan.

Mail : jean.feydy@ens.fr
Web : www.math.ens.fr/~feydy