Efficient and Scalable Spatial Regularization of Optimal Transport *Supplemental Material*

LUCAS BRIFAULT, Inria Saclay, France
DAVID COHEN-STEINER, Inria, Université Côte d'Azur, France
MATHIEU DESBRUN, Inria Saclay / Ecole Polytechnique, France

This document supplements our SIGGRAPH Asia 2025 paper by providing more figures and context on the foundations of our approach, as well as detailed computations and formulations to which the main text only briefly alludes.

A Background

We first recap the mathematical foundations upon which we build our approach. In particular, we review a number of basic terms (pushforward, μ-a.e., dual OT formulation, etc) used in the paper.

A.1 Primer on measure theory

Measure theory is concerned with measuring a notion of "size" (or equivalently, weight) for some subsets of a given set $\mathcal{X}$, generalizing the geometric notions of length, areas, etc. In general, we are interested in measuring subsets of $\mathcal{X}$ from a family $\Sigma_\mathcal{X} \subset 2^\mathcal{X}$ (a σ-algebra) that contains the empty set and is stable under complementary and countable union. The pair $(\mathcal{X}, \Sigma_\mathcal{X})$ is then called a *measurable space*, and an element of $\Sigma_\mathcal{X}$ is called a *measurable subset.*

Measurable maps. Given measurable spaces $(\mathcal{X}, \Sigma_\mathcal{X})$ and $(\mathcal{Y}, \Sigma_\mathcal{Y})$, a map $\phi : \mathcal{X} \to \mathcal{Y}$ is called *measurable* if it associates measurable subsets to measurable subsets, i.e., $\forall Y \in \Sigma_\mathcal{Y},\ \phi^{-1}(Y) \in \Sigma_\mathcal{X}$.

Measures. A measure μ is a weight assignment $\mu : \Sigma_\mathcal{X} \to \mathbb{R}_+ \cup \infty$ which satisfies that $\mu(\emptyset) = 0$ and, for any pairwise-disjoint countable family $\{X_i\}_{i\in\mathbb{N}}$, the measure of a union is the sum of its measures,

$$\mu\big(\bigcup_{i\in\mathbb{N}} X_i\big) \;=\; \sum_{i\in\mathbb{N}} \mu(X_i). \tag{1}$$

A measurable subset $X \subset \mathcal{X}$ is said to be negligible w.r.t μ if $\mu(X) = 0$, and a proposition $P(\cdot)$ on $\mathcal{X}$ is said to be true μ-almost everywhere (denoted as *μ-a.e.* in the paper) if the set $\{x \in \mathcal{X} | P(x) = \text{false}\}$ is negligible w.r.t. μ. Finally, the total mass of μ is the quantity $\mu(\mathcal{X}) \in \mathbb{R}_+ \cup \{\infty\}$; if it is equal to 1, then μ is a *probability measure.*

Integration. Once we have a measure μ on $\mathcal{X}$, we can formally define the Lebesgue integral of a non-negative measurable function $f : \mathcal{X} \to \mathbb{R}_+$ against μ in the following manner:

$$\int_\mathcal{X} f\, d\mu := \sup \Big\{ \sum_{i\in\mathbb{N}} \alpha_i \mu(X_i) \,\Big|\, \overbrace{\sum_{i\in\mathbb{N}} \alpha_i \mathbf{1}_{X_i}}^{\text{measurable}} \le f \quad \mu - \text{a.e.} \Big\}. \tag{2}$$

Integrable functions are simply measurable functions $f : \mathcal{X} \to \mathbb{R}$ such that $\int_\mathcal{X} |f|\, d\mu < \infty$, and we define their integrals via

$$\int_\mathcal{X} f d\mu := \int_\mathcal{X} \max(f, 0)\, d\mu \;-\; \int_\mathcal{X} \max(-f, 0)\, d\mu. \tag{3}$$

Push-forward. For any given measurable map $\phi : \mathcal{X} \to \mathcal{Y}$, the notion of *push-forward* specifies how a measure on $\mathcal{X}$ is naturally "transported" by ϕ onto a measure on $\mathcal{Y}$. Formally, the push-forward $\phi_\sharp\mu$ of a measure μ on $\Sigma_\mathcal{X}$ under ϕ is:

$$\forall Y \in \Sigma_\mathcal{Y}, \quad (\phi_\sharp\mu)(Y) := \mu\big(\phi^{-1}(Y)\big). \tag{4}$$

It generalizes the notion of change of variables in integration: if a function $g : \mathcal{Y} \to \mathbb{R}$ is integrable against $\phi_\sharp\mu$, then

$$\int_\mathcal{Y} g\, d(\phi_\sharp\mu) \;=\; \int_\mathcal{X} g \circ \phi\, d\mu. \tag{5}$$

Lebesgue spaces. For $p \in \mathbb{N}_{>0}$, consider the functional spaces of the form $\mathcal{L}^p(\mathcal{X};\mu) := \big\{ f : \mathcal{X} \to \mathbb{R}^d \text{ measurable} \big| \int_\mathcal{X} |f|^p d\mu < \infty \big\}$. Using the equivalence relation $\mathcal{R}$ such that $f\mathcal{R}g \iff f = g\ \mu - \text{a.e.}$, Lebesgue spaces $L^p(\mathcal{X};\mu)$ are defined as the quotient spaces $\mathcal{L}^p(\mathcal{X};\mu)/\mathcal{R}$ equipped with the norm $\|f\|_{L^p(\mathcal{X};\mu)} := \big(\int_\mathcal{X} |f|^p d\mu\big)^{1/p}$, making $L^p(\mathcal{X};\mu)$ a Banach space. For $p = 2$, the norm comes from the scalar product

$$\langle f, g\rangle_{L^2(\mathcal{X};\mu)} := \int_\mathcal{X} fg\, d\mu, \tag{6}$$

making $L^2(\mathcal{X};\mu)$ a Hilbert space.

Note that we can naturally extend this notion of Lebesgue spaces to functions with values in $\mathbb{R}^d$, defining in particular the space of interest $L^2(\mathcal{X}, \mathbb{R}^d; \mu)$, denoted $\mathcal{F}_\mathcal{X}$ for readability in the paper.

A.2 Basics of optimal transport

Given two measurable spaces $\mathcal{X}$ and $\mathcal{Y}$ and two probability measures, resp. μ and ν, on these spaces, we can talk about the cost of transporting one measure onto the other.

Monge problem. We can for instance try to minimize the Monge optimal transport (OT) cost $\int_\mathcal{X} c\big(x, T(x)\big) d\mu(x)$ (where $c(x, y)$ is a predefined cost of transporting "mass" from $x \in \mathcal{X}$ to $y \in \mathcal{Y}$) over *all* measurable maps $T : \mathcal{X} \to \mathcal{Y}$ such that $T_\sharp\mu = \nu$. Alas, there might not be such a map, and minimizing this cost is computationally difficult as the set of Monge maps do not form a convex set in general.

Kantorovitch extension. To overcome this issue, one may consider more general *transport plans* between μ and ν, which are probability measures π on $\mathcal{X}\times\mathcal{Y}$ with marginals μ on $\mathcal{X}$ and ν on $\mathcal{Y}$. The so-called Kantorovich optimal transport problem is a convex relaxation of the Monge problem, for which minimizers defined as

$$\operatorname*{arginf}_{\pi\in\Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{Y}} c(x, y)\, d\pi(x, y). \tag{7}$$

SA Conference Papers '25, December 15–18, 2025, Hong Kong, Hong Kong

ACM ISBN 979-8-4007-2137-3/2025/12
https://doi.org/10.1145/3757377.3763976

can be shown to exist under mild conditions. The minimum can also be found using the *dual* Kantorovich problem, formulated as

$$\sup_{\substack{\alpha\in L^1(\mathcal{X};\mu)\\ \beta\in L^1(\mathcal{Y};\nu)}} \int_{\mathcal{X}} \alpha \, d\mu + \int_{\mathcal{Y}} \beta \, d\nu \qquad \text{s.t.} \quad \alpha(x)+\beta(y) \le c(x,y) \quad \mu\otimes\nu-\text{a.e.} \tag{8}$$

Entropic regularization. Adding entropic regularization to the objective function of the primal OT problem turns it into

$$\inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{Y}} c(x,y)\, d\pi(x,y) \; + \; \epsilon H(\pi) \tag{9}$$

where $H(\pi) = \mathrm{KL}(\pi|\mu\otimes\nu)$ (where KL means Kullback–Leibler divergence), and the dual problem now involves a "soft" penalty replacing the inequality from Eq. (8), allowing faster solvers such as Sinkhorn iterations [Cuturi 2013].

Typical transport costs. The choice of transport cost $c(x,y)$ in Eq. (7) is key to establishing a useful transport plan for a given application. We can leverage particular (linear combinations of) transport costs to promote a particular behavior of the resulting optimal transport plan. In many cases, the *Euclidean cost function* $c(x,y)=|x-y|^2$ is often inadequate because of its high dependency on (near) isometries: it relies too much on the given embeddings and not enough on the intrinsic geometry of the objects. The knowledge of k pointwise landmarks $(x_1,\cdots,x_k)\in\mathcal{X}^k$ and their correspondences $(y_1,\cdots,y_K)\in\mathcal{Y}^k$, available in many applications, can instead be exploited to derive an intrinsic transport cost:

$$c(x,y) = \sum_{k=1}^{K} \left|\theta\big(d_{\mathcal{X}}(x_k,x)\big) - \theta\big(d_{\mathcal{Y}}(y_k,y)\big)\right|^2, \tag{10}$$

where $\theta:\mathbb{R}_{>0}\to\mathbb{R}$ is any continuous monotonous function while $d_{\mathcal{X}}$ and $d_{\mathcal{Y}}$ are the geodesic distances of $\mathcal{X}$ and $\mathcal{Y}$ respectively. In applications where plans should optimally be with isometric maps (or more generally, maps with specific geometric properties), local shape (feature) descriptors that are invariant to isometries can also be leveraged to design transport costs: if descriptors $F:\mathcal{X}\to\mathbb{R}^N$ and $G:\mathcal{Y}\to\mathbb{R}^N$ are such that if there exists an isometry $\psi:\mathcal{X}\to\mathcal{Y}$ with $\psi_\sharp\mu=\nu$, then $F=G\circ\psi$, the cost:

$$c(x,y) = \left\|F(x)-G(y)\right\|^2 \tag{11}$$

can help steer the optimal plan to have the intended property — where typical feature descriptors F and G may include the Heat Kernel Signature [Sun et al. 2009], diffusion maps [Coifman et al. 2005], or any other isometry-invariant signature.

B About variance loss

For a plan $\pi\in\Pi(\mu,\nu)$, let $g\in\mathcal{F}_{\mathcal{Y}}$ and $\phi=\overrightarrow{m}_\pi\in\mathcal{F}_{\mathcal{X}}$. One has:

$$\begin{aligned}
\mathrm{Var}_\nu &:= \int_{\mathcal{Y}} \left|y-\bar{\nu}\right|^2 d\nu(y) = \int_{\mathcal{X}\times\mathcal{Y}} \Big|y-\phi(x)+\phi(x)-\underbrace{\bar{\nu}}_{=\mathbb{E}_\mu[\phi]}\Big|^2 d\pi(x,y)\\
&= \int_{\mathcal{X}\times\mathcal{Y}} \left|y-\phi(x)\right|^2 d\pi(x,y)\\
&\quad + \underbrace{\int_{\mathcal{X}} \left|\phi(x)-\mathbb{E}_\mu[\phi]\right|^2 d\mu(x)}_{=\mathrm{Var}_\mu[\phi]}\\
&\quad - \int_{\mathcal{X}\times\mathcal{Y}} \big(y-\phi(x)\big)\cdot\big(\phi(x)-\mathbb{E}_\mu[\phi]\big)\, d\pi(x,y).
\end{aligned} \tag{12}$$

This last integral vanishes since, by definition of ϕ,

$$\int_{\mathcal{X}\times\mathcal{Y}} y\cdot\big(\phi(x)-\mathbb{E}_\mu[\phi]\big)\, d\pi(x,y) = \underbrace{\int_{\mathcal{X}} \phi\,\big(\phi-\mathbb{E}_\mu[\phi]\big)\, d\mu}_{=\int_{\mathcal{X}\times\mathcal{Y}}\phi(x)\cdot\left(\phi(x)-\mathbb{E}_\mu[\phi]\right)\, d\pi(x,y)}$$

and of course $\int_{\mathcal{X}} F(x)\, d\mu(x) = \int_{\mathcal{X}\times\mathcal{Y}} F(x)\, d\pi(x,y)$. Therefore, one has that $\mathrm{Var}_\nu = \mathrm{Var}_\mu[\phi] + \int_{\mathcal{X}\times\mathcal{Y}} \left|\phi(x)-y\right|^2 d\pi(x,y)$. As a consequence, $\mathrm{Var}_\nu \ge \mathrm{Var}_\mu[\phi]$'. Now, if equality holds between Var_ν and $\mathrm{Var}_\mu[\phi]$, then we can deduce that

$$\begin{aligned}
0 &\le W_2\big(\phi_\sharp\mu,\ \nu\big)^2\\
&= \inf_{\pi'\in\Pi(\phi_\sharp\mu,\nu)} \int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^2\, d\pi'(x,y)\\
&= \inf_{\pi"\in\Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{Y}} |\phi(x)-y|^2\, d\pi"(x,y)\\
&\le \int_{\mathcal{X}\times\mathcal{Y}} |\phi(x)-y|^2\, d\pi(x,y) = 0,
\end{aligned} \tag{13}$$

i.e., $W_2\big(\phi_\sharp\mu,\ \nu\big)=0$ — which means that $\phi_\sharp\mu=\nu$. Finally, note the converse is true as well trivially: if $\phi_\sharp\mu=\nu$, then $\mathrm{Var}_\nu=\mathrm{Var}_\mu[\phi]$.

C Gradient expression for $\widehat{Q}$

Using $\lambda(\phi)=-\mathbf{A}\phi-b$, we have

$$\begin{aligned}
\widehat{Q}(\phi) &= \tfrac{1}{2}\left\langle \phi, \mathbf{A}\phi\right\rangle_{\mathcal{F}_{\mathcal{X}}} - E_\epsilon^{OT}\big(\lambda(\phi)\big)\\
&= \tfrac{1}{2}\left\langle \phi, \mathbf{A}\phi\right\rangle_{\mathcal{F}_{\mathcal{X}}} - \inf_{\pi\in\Pi(\mu,\nu)} \Bigg[\underbrace{\int c_{\lambda(\phi)}\, d\pi \; + \; \epsilon.H(\pi)}_{F(\pi,\phi)}\Bigg]
\end{aligned} \tag{14}$$

where $c_\lambda(x,y) = c(x,y) - \lambda(x)\cdot g(y)$. The gradient of the first term is clearly $\mathbf{A}\phi$. For the second term, remark that since F is continuous and since the problem $min_\pi F(\phi,\pi)$ has a unique solution $\pi[\phi]$, Danskin's theorem implies $\nabla\big(\inf_\pi F(\phi,\pi)\big) = \nabla_\phi F\big(\phi,\pi[\phi]\big)$.

Noting that

$$\begin{aligned}
\nabla_\phi F(\phi,\pi) &= \nabla_\phi \int (\mathbf{A}\phi)(x)\cdot g(y)\, d\pi(x,y)\\
&= \nabla_\phi \left\langle \mathbf{A}\phi, \overrightarrow{m}_\pi\right\rangle_{\mathcal{F}_{\mathcal{X}}} = \mathbf{A}\overrightarrow{m}_\pi
\end{aligned} \tag{15}$$

where the last equality holds by symmetry of $\mathbf{A}$, we end up with:

$$\nabla\widehat{Q}(\phi) \;=\; \mathbf{A}\phi - \mathbf{A}\pi[\phi]^* g. \tag{16}$$

D Discrete optimal transport

In this section, we provide details on how optimal transport of surface meshes is achieved in practice in our OT context.

Consider two discrete probability measures μ and ν in $\mathbb{R}^d$ with

$$\mu \;:=\; \sum_{i=1}^{M} \mu_i\, \delta_{x_i}, \qquad \nu \;:=\; \sum_{j=1}^{N} \nu_j\, \delta_{y_j}, \tag{17}$$

where μ_i, ν_j are positive weights such that $\sum_i \mu_i = \sum_j \nu_j = 1$, δ_x represents a Dirac mass at a point $x\in\mathbb{R}^d$, and $(x_1,\cdots,x_M)\in\mathbb{R}^{d\times M}$, $(y_1,\cdots,y_N)\in\mathbb{R}^{d\times N}$ are two sets of points in $\mathbb{R}^d$.

In our examples, such discrete measures can be derived from surface meshes, where the $\{x_i\}_i$ (and/or $\{y_j\}_j$) are the vertex positions of

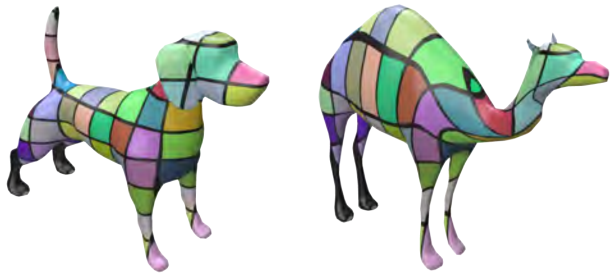

Fig. 1. **Texture for Dog to camel Example.** A source texture (left) on the dog is pushed-forward (transferred) by the forward mean map (right) shown in the teaser of the main text.

the mesh, and the associated weights are given by the normalized 1-ring area of each vertex, that is:

$$\mu_i \coloneqq Z^{-1} \sum_{\tau \in T | x_i \in \tau} area(\tau), \quad \text{with } Z = 3 \sum_{\tau \in T} area(\tau), \tag{18}$$

where we used T to refer to the set of faces of the triangle mesh. The discrete measures can also derive from regular grids, in which case we use $\mu_i = \frac{1}{M}$ (resp. $\nu_j = \frac{1}{N}$) and the $(x_i)_i$ (and $(y_j)_j$) are the centers of the cells.

The entropic optimal transport problem between μ and ν is then

$$\begin{array}{rrcll} \inf_{\mathbf{\Pi} \in \mathbb{R}^{M\times N}} & \mathbf{C} : \mathbf{\Pi} & + & \epsilon\, H(\mathbf{\Pi}) & \\ \text{s.t.} & \sum_j \Pi_{ij} & = & \mu_i & \forall\, 1 \le i \le M \\ & \sum_i \Pi_{ij} & = & \nu_j & \forall\, 1 \le j \le N \end{array} \tag{19}$$

where $\mathbf{C}$ is a $M \times N$ matrix with coefficients $C_{ij} = c(x_i, y_j)$ (for some cost function c), the “:” symbol denotes the Frobenius scalar product, $H(\mathbf{\Pi}) \coloneqq \sum_{ij} \Pi_{ij}\big(\log(\Pi_{ij}) - 1\big)$, and $\epsilon > 0$. Our method requires to solve this kind of problem multiple times with various cost functions $c(.,.)$, for which we rely on Sinkhorn's algorithm and its stable and efficient variants based on log-domain, symmetrization and annealing [see Feydy 2020, Alg. 3.5], that we implemented on the GPU for a much faster computation, as it parallelizes well.

E Practical Hessian-based energy for surfaces

In this section, we provide a more in-depth derivation of our Hessian-based energy function for triangle meshes.

E.1 Continuous case

Let $\mathcal{X} \subset \mathbb{R}^3$ be a compact smooth surface patch and $\varphi : \mathcal{X} \longrightarrow \mathbb{R}$ be s smooth function defined over this patch. The Hessian operator $\mathbf{H}[\cdot]$ on functions of $\mathcal{X}$ returning at each point $x \in \mathcal{X}$ a 3×3 matrix can be defined through the Riemannian gradient operator $\nabla_{\mathcal{X}}$ on $\mathcal{X}$ via

$$\forall x \in \mathcal{X}, \quad \mathbf{H}[\varphi](x) \coloneqq \nabla_{\mathcal{X}}(\nabla_{\mathcal{X}}\varphi)(x)\ \mathbf{P}_{\mathcal{X}}(x), \tag{20}$$

where $\mathbf{P}_{\mathcal{X}}(x)$ denotes the orthogonal projection operator on the tangent plane $T_x\mathcal{X}$ at x. Note that $\nabla_{\mathcal{X}} u$ is to be understood coordinate by coordinate if u has values in $\mathbb{R}^d$; i.e., $\nabla_{\mathcal{X}}\varphi(x)$ returns a 3D vector in $T_x\mathcal{X}$, while $\nabla_{\mathcal{X}}(\nabla_{\mathcal{X}}\varphi)(x)$ returns a 3×3 matrix, explaining the final presence of the orthogonal projection operator $\mathbf{P}_{\mathcal{X}}(x)$ to remove the components normal to $\mathcal{X}$ at x in the rows of the matrix.

Note that functions of the form $\varphi(x) = a \cdot x + b$ are not, in general, in the kernel of $\mathbf{H}$: we only compute tangent variations of φ since it is only defined on $\mathcal{X}$. If we had access to a smooth "extension" $\tilde{\varphi} : \mathbb{R}^3 \longrightarrow \mathbb{R}$ (still satisfying $\tilde{\varphi}(x) = \varphi(x)\ \forall x \in \mathcal{X}$), we could simply compute $\nabla^2\tilde{\varphi}$ everywhere on $\mathcal{X}$, with ∇ being the gradient in $\mathbb{R}^3$. Usually, no such extension is given, but since in the end we are interested in the energy $\left\|\nabla^2\tilde{\varphi}\right\|^2_{\mathcal{F}_{\mathcal{X}}}$, we might as well choose an extension that minimizes this energy. Now, for any $x \in \mathcal{X}$, $\nabla^2\tilde{\varphi}(x)$ is a 3×3 matrix that can be decomposed into four orthogonal parts as follows:

$$\begin{aligned} \nabla^2\tilde{\varphi}(x) = \mathbf{P}_{\mathcal{X}}(x)\nabla^2\tilde{\varphi}(x)\mathbf{P}_{\mathcal{X}}(x) \\ + \mathbf{P}_{\mathcal{X}}(x)\nabla^2\tilde{\varphi}(x)(I - \mathbf{P}_{\mathcal{X}}(x)) \\ + (I - \mathbf{P}_{\mathcal{X}}(x))\nabla^2\tilde{\varphi}(x)\mathbf{P}_{\mathcal{X}}(x) \\ + (I - \mathbf{P}_{\mathcal{X}}(x))\nabla^2\tilde{\varphi}(x)(I - \mathbf{P}_{\mathcal{X}}(x)), \end{aligned} \tag{21}$$

and the squared Frobenius norm of $\nabla^2\tilde{\varphi}$ is simply the sum of the squared Frobenius norm of each of these four terms. But it is always possible to choose a $\tilde{\varphi}$ such that the last term vanishes for all $x \in \mathcal{X}$ without affecting the other components. Note also that, since $\nabla^2\tilde{\varphi}(x)$ is always symmetric, the second and the third term are the same up to a transposition. Now, since $\nabla_{\mathcal{X}} u = \mathbf{P}_{\mathcal{X}}(\nabla u)$ for $u : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ and that $I - \mathbf{P}_{\mathcal{X}}$ can actually be rewritten $\boldsymbol{n} \otimes \boldsymbol{n}$ (where $\mathbf{n} : \mathcal{X} \longrightarrow \mathbb{R}^3$ denotes the Gauss map of $\mathcal{X}$), we just need to minimize $\|\nabla_{\mathcal{X}}\nabla\tilde{\varphi}\|^2_{\mathcal{F}_{\mathcal{X}}} + \|(\nabla_{\mathcal{X}}\nabla\tilde{\varphi}) \cdot \boldsymbol{n}\|^2_{\mathcal{F}_{\mathcal{X}}}$.

Since we do not want to build $\tilde{\varphi}$ over the whole $\mathbb{R}^3$, we make a change of variable using $v = \nabla\tilde{\varphi}$, where $v : \mathcal{X} \longrightarrow \mathbb{R}^3$ will have to satisfy the constraint $\mathbf{P}_{\mathcal{X}}(x)v(x) = \nabla_{\mathcal{X}}\varphi(x)$, which is sufficient for our needs. The Hessian matrix for the optimal $\tilde{\varphi}$ is then given by ∇v, where v is a solution of:

$$\inf\left\{\|\nabla_{\mathcal{X}} v\|^2_{\mathcal{F}_{\mathcal{X}}} + \|(\nabla_{\mathcal{X}} v)\cdot\boldsymbol{n}\|^2_{\mathcal{F}_{\mathcal{X}}} \,\middle|\, v \in \mathcal{F}_{\mathcal{X}}, \mathbf{P}_{\mathcal{X}}(x)v(x) = \nabla_{\mathcal{X}}\varphi(x)\ \forall x \in \mathcal{X}\right\}.$$

The value of this infimum is precisely the energy $e(\varphi)$ defined in Eq. (27) of the main paper. Indeed, the constraint can even be made more explicit by writing the decomposition $v(x) = \nabla_{\mathcal{X}}\varphi(x) - h(x)\mathbf{n}(x)$, for some function $h \in \mathcal{F}_{\mathcal{X}}$ which is the only remaining free variable. This leads us to $e(\varphi) = \inf_{h \in \mathcal{F}_{\mathcal{X}}} e(\varphi, h)$ with:

$$e(\varphi, h) = \|\nabla_{\mathcal{X}}(\nabla_{\mathcal{X}}\varphi + h\boldsymbol{n})\|^2_{\mathcal{F}_{\mathcal{X}}} + \|(\nabla_{\mathcal{X}}\varphi + h\boldsymbol{n}) \cdot \boldsymbol{n}\|^2_{\mathcal{F}_{\mathcal{X}}}\,. \tag{22}$$

Notice that by construction, the minimizers of this energy are exactly the affine functions of $\mathbb{R}^3$, i.e. $\{x \in \mathcal{X} \mapsto a \cdot x + b \mid a, b \in \mathbb{R}^3\}$.

E.2 Implementation on triangle meshes

Consider a triangle mesh (V, E, T) in $\mathbb{R}^3$ defined by vertices V, edges E and triangles T. We use the space $\mathcal{F}_V$ of continuous functions that are affine on each triangle, i.e.,

$$\mathcal{F}_V \coloneqq \left\{\sum_{i \in V} f_i\ \varphi_i \,\middle|\, f_i \in \mathbb{R}\right\} \tag{23}$$

where the set $\{\varphi_i\}_{i \in V}$ contains the usual 'hat' basis functions on our mesh, that is, the family of functions on the surface mesh that are linear inside triangles and satisfy $\varphi_i(x_j) = \delta_{ij}, \forall i, j \in V$. We also use the space of functions that are constant per triangle

$$\mathcal{F}_T \ \coloneqq\ \left\{\sum_{\tau \in T} f_\tau\, \mathbf{1}_\tau \,\middle|\, f_\tau \in \mathbb{R}\right\}, \tag{24}$$

so that the gradient operator on $\mathcal{F}_V$ is expressed as:

$$\nabla_{V\to T} : \left\{ \begin{array}{ccc} \mathcal{F}_V & \longrightarrow & \mathcal{F}_T^3 \\ \sum_i f_i\ \varphi_i & \longmapsto & \sum_{\tau \in T}\big(\sum_{i \in V} f_i \nabla\varphi_i(c_\tau)\big)\ \mathbf{1}_\tau \end{array} \right. \tag{25}$$

where c_τ is the barycenter of triangle τ. Now, because we have an operator $\nabla_{V\to T} : \mathcal{F}_V \to \mathcal{F}_T$ between two Euclidean spaces (using the scalar product induced by L^2), the adjoint of the gradient operator provides us with a discrete divergence operator $\mathrm{div}_{T\to V} = -\left(\nabla_{V\to T}\right)^*$. More precisely:

$$\mathrm{div}_{T\to V} : \left\{ \begin{array}{ccc} \mathcal{F}_T^3 & \longrightarrow & \mathcal{F}_V \\ \sum_\tau v_\tau\, \mathbf{1}_\tau & \longmapsto & -\sum_{i\in V} \left(\mathbf{M}^{-1} Dv\right)_i \varphi_i \end{array} \right. \tag{26}$$

with the linear operator $D : \mathcal{F}_T^3 \to \mathbb{R}^{|V|}$ defined such that $(Dv)_i = \sum_{\tau\in T} |\tau| \nabla\varphi_i^\tau \cdot v_\tau$ (with $|\tau|$ being the area of the triangle τ), and $\mathbf{M}$ being the $|V| \times |V|$ "mass " matrix with coefficients $M_{ij} = \int \varphi_i \varphi_j dA$. For any smooth function f on a smooth surface $\mathcal{X}$, note that one can always write $\nabla_{\mathcal{X}} f = \mathrm{div}_{\mathcal{X}}\left(f\, \mathbf{I}_3\right)$. Mimicking this identity, we define another gradient operator on triangulated surfaces as:

$$\nabla_{T\to V} : \left| \begin{array}{ccc} \mathcal{F}_T & \longrightarrow & \mathcal{F}_V^3 \\ f & \longmapsto & \mathrm{div}_{T\to V}\left(f\, \mathbf{I}_3\right) \end{array} \right. \tag{27}$$

(Here $\mathrm{div}_{\mathcal{X}}$ and $\mathrm{div}_{T\to V}$ are to be understood as column-wise operators when applied to matrix fields on smooth vs. discrete surfaces.) Those operators make us able to define a discrete version of the energy (22): for functions $\varphi \in \mathcal{F}_V$ and $h \in \mathcal{F}_T$, we simply define

$$\hat{e}(\varphi, h) \;\coloneqq\; \left\|\nabla_{T\to V}\left(\nabla_{V\to T}\varphi + h\boldsymbol{n}\right)\right\|_{\mathcal{F}_V}^2 + \left\|\left(\nabla_{V\to T}\varphi + h\boldsymbol{n}\right)\cdot \boldsymbol{n}\right\|_{\mathcal{F}_T}^2 \tag{28}$$

where $\nabla_{T\to V}$ is applied coordinate by coordinate.

F Other Relevant Details

We finally review a few more technical explanations that the main text omits for conciseness. Fig. 1 also provides a visualization of texture transfer from the example used in the main text.

F.1 Our LDDMM implementation from Fig. 5

At the core of the LDDMM (Large Deformation Diffeomorphic Metric Mapping) framework lies a deformation model based on the flow $\phi^v : \mathbb{R}^d \to \mathbb{R}^d$ of vector fields $v : [0,1]\times\mathbb{R}^d \to \mathbb{R}^d$ in the ambient space $\mathbb{R}^d$ (typically, $d = 2$ or 3). In our case, a deformation ϕ^v is applied to a source mesh S, producing the mesh S^{ϕ^v} by simply moving every vertex x_i of S to the position $\phi^v(x_i)$. Furthermore, the vector fields that we consider are of the form:

$$v(x) \;=\; \sum_i K_\sigma(x, c_i)\, p_i, \tag{29}$$

with $p_i \in \mathbb{R}^2$, $K_\sigma(x,y) = \exp\left(-|x-y|^2/(2\sigma^2)\right)$, and the c_i's are points in $\mathbb{R}^2$ on a grid around both the source and target meshes S and T.

In our LDDMM example from Fig. 5, we use a typical regularization cost for the deformation induced by vector fields v of the form given in Eq. (29) that is expressed as:

$$\mathrm{Reg}(v) \coloneqq \sum_{ij} K_\sigma(c_i, c_j)\, p_i \cdot p_j \;. \tag{30}$$

For the data attachment term between the deformed source mesh S^{ϕ^v} and the target mesh T (with vertices $(y_j)_j \in \mathbb{R}^2$), we use:

$$\mathrm{data}(S^{\phi^v}, T) \;=\; W_\epsilon^2\left(\sum_i \mu_i \delta_{\phi^v(x_i)}\;,\; \sum_j \nu_j \delta_{y_j}\right), \tag{31}$$

where μ_i's and ν_j's are the weights associated to the meshes S^{ϕ^v} and T respectively according to Eq. (18), as well as a squared distance term on the corresponding landmarks $\mathrm{land}(\phi^v) \coloneqq \sum_{k=1}^3 |\phi^v(x_k^l) - y_k^l|^2$ (where $\{x_k^l\}_k$ and $\{y_k^l\}_k$ denote the respective landmark positions on the source and target meshes in $\mathbb{R}^2$). The resulting deformation ϕ^v is computed by minimizing the total energy $\mathrm{Reg}(v) + \alpha\, \mathrm{data}\left(S^{\phi^v}, T\right) + \beta\, \mathrm{land}(\phi^v)$. We tried various combinations of parameters $\sigma, \alpha, \beta > 0$ and grid size for the $\{c_i\}_i$, and we used the best looking result in the figure.

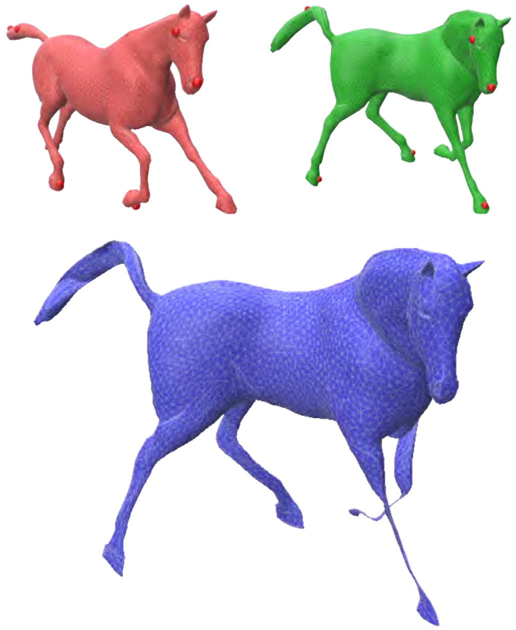

Fig. 2. **Failure case of the purely convex formulation.** An optimal mean map (blue) between a source (red) and target (green) poses of a horse uing landmark-based distances can fail when computed with our purely convex Hessian-regularized formulation: the front legs, and partially the tail, collapse on themselves because of variance loss.

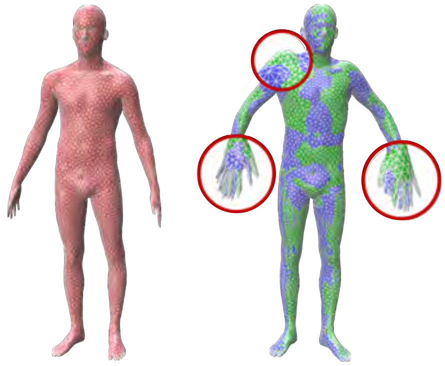

Fig. 3. **Another failure case of the purely convex formulation.** The optimal forward mean map (blue) between a source (red) and target (green) poses of a human using Euclidean distance as the transport cost is overall ok, but small details, like the fingers or the tip of the right shoulder are not well recovered (see circled zoomed-in views).

F.2 Examples of failure with the convex formulation

Our convex spatially regularized OT method (without Mongification) works well when the cost function $c(.,.)$ (in Eq. (6) of the main text) already provides a good hint for which points of the source and target measures should be matched. For instance, a Euclidean cost works well if the source and target shapes are in similar poses, like

Table 1. **Size and timings of our tests**: besides the number of vertices in the source and target meshes, we indicate if Mongification was used, the total number of Sinkhorn calls used, and the total computational time.

Fig.	#vert. source	#vert. target	Mong.	# Sink.	Time
1 & 13	2487	2464	No	3	5”
4 & 11	1502	2502	No	3	6”
5	561	845	No	50	4”
6 & 7	5901	6332	Yes	20	1’57”
8	3087	4466	Yes	20	1’01”
9 & 10	6198	6021	Yes	20	1’55”
12	5070	4934	Yes	30	1’43”
14	2930	5258	Yes	40	59”

in Fig 1 or 4 of the main text. In some cases though, it is hard to find such a good cost function, and even the use of cost functions based on landmarks may have trouble distinguishing different points of each shape if there are not enough landmarks: if this happens for large enough regions, then even regularization is often not enough to prevent the mean map from “collapsing”. This phenomenon is illustrated in Fig. 2 where we used a landark-based transport cost function from Eq. (10) with 7 landmarks: two of the legs of the horse end up shrinking in diameter. Fig. 3 in the main text shows that Mongification fixes this issue, even if only three landmarks are used.

Even when the Euclidean cost function is seemingly well suited to match a pair of surfaces, like in Fig.3 where the poses are not vastly different, using the purely convex method provides a good registration for most of the shape, but some details are off: the mean map is not guaranteed to land on the target surface, in fact, failing for the parts with more detailed structure (like the fingers of both hands) without Mongification.

F.3 Computational times

For the numerical examples in the main paper (identified with their respective figure number), Tab. 1 specifies the size of each source and target mesh through their numbers of vertices, whether Mongification was used or not, the total number of calls to Sinkhorn’s algorithm before convergence, and the computational (wall clock) time. Results, taking from seconds up to two minutes, were computed on an Intel Xeon based computer with 64 GB RAM, equipped with an Nvidia RTX 6000 GPU card with 48 GB VRAM.

References

Ronald R. Coifman, Stephane Lafon, Ann B. Lee, Mauro Maggioni, Boaz Nadler, Frederick Warner, and Steven W. Zucker. 2005. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *Proceedings of the National Academy of Sciences* 102, 21 (2005), 7426–7431.

Marco Cuturi. 2013. Sinkhorn distances: Lightspeed computation of optimal transport. *NeurIPS* 26 (2013), 2292–2300.

Jean Feydy. 2020. *Geometric data analysis, beyond convolutions.* Ph. D. Dissertation. Université Paris-Saclay.

Jian Sun, Maks Ovsjanikov, and Leonidas Guibas. 2009. A concise and provably informative multi-scale signature based on heat diffusion. In *Comp. Graph. Forum*, Vol. 28. 1383–1392.