Mapping Estimation for Discrete Optimal Transport Supplementary Material

Michaël Perrot
Univ Lyon, UJM-Saint-Etienne, CNRS,
Lab. Hubert Curien UMR 5516, F-42023
`michael.perrot@univ-st-etienne.fr`

Nicolas Courty
Université de Bretagne Sud,
IRISA, UMR 6074, CNRS,
`courty@univ-ubs.fr`

Rémi Flamary
Université Côte d'Azur,
Lagrange, UMR 7293 , CNRS, OCA
`remi.flamary@unice.fr`

Amaury Habrard
Univ Lyon, UJM-Saint-Etienne, CNRS,
Lab. Hubert Curien UMR 5516, F-42023
`amaury.habrard@univ-st-etienne.fr`

1 Proof of the Joint Convexity of the Optimization Problem

We recall the optimization problem:

$$\operatorname*{arg\,min}_{T\in\mathcal{H},\gamma\in\hat{\Pi}} f(\gamma,T)=\frac{1}{n_s d_t}\left\|T(\mathbf{X}_s)-n_s\gamma\mathbf{X}_t\right\|_{\mathcal{F}}^2+\frac{\lambda_\gamma}{\max(\mathbf{C})}\left\langle\gamma,\mathbf{C}\right\rangle_{\mathcal{F}}+\frac{\lambda_T}{d_s d_t}R(T) \tag{1}$$

and the theorem:

Theorem 1. *Let $\mathcal{H}$ be a convex space and $R(\cdot)$ be a convex function. Problem* (1) *is jointly convex in T and γ.*

Proof. First of all recall that a sum of jointly convex functions is jointly convex. Hence it is sufficient to show that the three terms of optimization problem (1) are jointly convex. We note $f_1(\gamma,T)=\frac{1}{n_s d_t}\left\|T(\mathbf{X}_s)-n_s\gamma\mathbf{X}_t\right\|_{\mathcal{F}}^2$, $f_2(\gamma)=\frac{\lambda_\gamma}{\max(\mathbf{C})}\left\langle\gamma,\mathbf{C}\right\rangle_{\mathcal{F}}$ and $f_3(T)=\frac{\lambda_T}{d_s d_t}R(T)$. f_1 depends on both T and γ and controls the proximity between the transformation induced by T and the barycentric interpolation obtained from γ. f_2 only depends on γ, it corresponds to the standard term minimized to solve the optimal transport problem. f_3 regularizes T to ensure a better generalization.

Note that f_2 and f_3 are by construction jointly convex in γ and T. We will show that the f_1 is also jointly convex. Let $g(\gamma,T)=\left\|T(\mathbf{X}_s)-n_s\gamma\mathbf{X}_t\right\|_{\mathcal{F}}$, we want to show that:

$$g(t\gamma_1+(t-1)\gamma_2,tT_1+(1-t)T_2)\leq tg(\gamma_1,T_1)+(1-t)g(\gamma_2,T_2).$$

We have:

$$\begin{aligned}&\left\|(tT_1+(1-t)T_2)(\mathbf{X}_s)-n_s(t\gamma_1+(t-1)\gamma_2)\mathbf{X}_t\right\|_{\mathcal{F}}\\ &\qquad\qquad\qquad\qquad\text{(Triangle inequality and definition of }\mathcal{H}.)\\ &\quad\leq\left\|tT_1(\mathbf{X}_s)-tn_s\gamma_1\mathbf{X}_t\right\|_{\mathcal{F}}+\left\|(1-t)T_2(\mathbf{X}_s)-(1-t)n_s\gamma_2\mathbf{X}_t\right\|_{\mathcal{F}}\\ &\qquad\qquad\qquad\qquad(t\in[0,1].)\\ &\quad=t\left\|T_1(\mathbf{X}_s)-n_s\gamma_1\mathbf{X}_t\right\|_{\mathcal{F}}+(1-t)\left\|T_2(\mathbf{X}_s)-n_s\gamma_2\mathbf{X}_t\right\|_{\mathcal{F}}\end{aligned}$$

30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain.

Furthermore noting that g is convex and positive we have:

$$\begin{aligned}
&\left[g(t\gamma_1 + (t-1)\gamma_2, tT_1 + (1-t)T_2)\right]^2 \\
&\qquad\qquad\qquad\qquad (\forall x \in \mathbb{R}^+, x \to x^2 \text{ is non decreasing.}) \\
&\leq \left[tg(\gamma_1, T_1) + (1-t)g(\gamma_2, T_2)\right]^2 \\
&\qquad\qquad\qquad\qquad (\forall x \in \mathbb{R}, x \to x^2 \text{ is convex.}) \\
&\leq t\left[g(\gamma_1, T_1)\right]^2 + (1-t)\left[g(\gamma_2, T_2)\right]^2 .
\end{aligned}$$

Noting that $f_1(\gamma, T) = \frac{1}{n_s d_t} g(\gamma, T)^2$ concludes the proof. $\square$

2 Details about Block Coordinate Descent

To solve optimization problem (1) we propose to use a block-coordinate descent approach. As such we need to find an efficient way to solve for γ when T is fixed and to solve for T when γ is fixed.

Solving for γ with T fixed In this case we want to solve:

$$\arg\min_{\gamma \in \hat{\Pi}} f(\gamma, T) = \frac{1}{n_s d_t} \left\| T(\mathbf{X}_s) - n_s \gamma \mathbf{X}_t \right\|_{\mathcal{F}}^2 + \frac{\lambda_\gamma}{\max(\mathbf{C})} \langle \gamma, \mathbf{C} \rangle_{\mathcal{F}} + \frac{\lambda_T}{d_s d_t} R(T) \tag{2}$$

where T is the current transformation. To solve such an optimization problem a common approach is to use the Frank-Wolfe algorithm [1, 2]. It is a procedure for solving any convex constrained optimization problems with a convex and continuously differentiable objective function over a compact convex subset of any vector space. This algorithm can find an ϵ approximation of the optimal solution in $O(1/\epsilon)$ iterations [3]. A detailed algorithm is given in Section 3.

Solving for T with γ fixed In this case we want to solve:

$$\arg\min_{T \in \mathcal{H}} f(\gamma, T) = \frac{1}{n_s d_t} \left\| T(\mathbf{X}_s) - n_s \gamma \mathbf{X}_t \right\|_{\mathcal{F}}^2 + \frac{\lambda_\gamma}{\max(\mathbf{C})} \langle \gamma, \mathbf{C} \rangle_{\mathcal{F}} + \frac{\lambda_T}{d_s d_t} R(T) \tag{3}$$

where γ is the current mapping between the examples. The solution to this optimization problem depends on the form of $\mathcal{H}$ and R. This is discussed in detail in Section 3.2 in the main paper.

3 Detailed Frank-Wolfe algorithm

We propose in Algorithm 1 a detailed version of the Frank-Wolfe approach for solving problem (2).

Algorithm 1: Updating γ with the Frank-Wolfe algorithm.

```
input  : The current values of $\gamma$ and $T$.
output : The new value of $\gamma$.
begin
    Initialize $k = 0$ and $\gamma^0 = \gamma$
    repeat
        Solve $\mathbf{S}^k = \arg\min_{\mathbf{S} \in \hat{\Pi}} \left\langle \mathbf{S}, \nabla f(\gamma^k, T) \right\rangle_{\mathcal{F}}$ with
            $\nabla f(\gamma, T) = \frac{\lambda_\gamma}{\max(\mathbf{C})}\mathbf{C} - \frac{2}{n_s d_t} n_s T(\mathbf{X}_s)\mathbf{X}_t^T + \frac{2}{n_s d_t} n_s^2 \gamma \mathbf{X}_t \mathbf{X}_t^T$.
        Find the optimal step $\alpha^k$ satisfying the Armijo rule that minimizes $f\left((1-\alpha)\gamma^k + \alpha\mathbf{S}^k, T\right)$.
        Update $\gamma^{k+1} = (1-\alpha)\gamma^k + \alpha\mathbf{S}^k$ and $k = k + 1$.
    until convergence
```

4 Bias including version of $\mathcal{H}$

We present the bias including version of $\mathcal{H}$ both in the linear and the non-linear case.

Biased linear transformations In the biased linear case we have:

$$\mathcal{H} = \left\{ T : \exists\, \mathbf{L} \in \mathbb{R}^{d^s \times d^t}, \exists\, \mathbf{b} \in \mathbb{R}^{d^t}, \forall \mathbf{x} \in \Omega_{\mathcal{S}}, T(\mathbf{x}^s) = \mathbf{x}^{sT}\mathbf{L} + \mathbf{b}^T = \begin{pmatrix} \mathbf{x}^{sT} & 1 \end{pmatrix} \begin{pmatrix} \mathbf{L} \\ \mathbf{b}^T \end{pmatrix} \right\}. \tag{4}$$

In this case, optimization problem 1 becomes:

$$\underset{\begin{pmatrix} \mathbf{L} \\ \mathbf{b}^T \end{pmatrix} \in \mathbb{R}^{d^s+1 \times d^t}, \gamma \in \hat{\Pi}}{\arg\min} \frac{1}{n_s d_t} \left\| \begin{pmatrix} \mathbf{X}_s & \mathbf{1} \end{pmatrix} \begin{pmatrix} \mathbf{L} \\ \mathbf{b}^T \end{pmatrix} - n_s \gamma \mathbf{X}_t \right\|_{\mathcal{F}}^2 + \frac{\lambda_\gamma}{\max(\mathbf{C})} \langle \gamma, \mathbf{C} \rangle_{\mathcal{F}} + \frac{\lambda_T}{d_s d_t} \|\mathbf{L} - \mathbf{I}\|_{\mathcal{F}}^2 . \tag{5}$$

As in the non biased case, it is possible to find a closed form solution for $\begin{pmatrix} \mathbf{L} \\ \mathbf{b}^T \end{pmatrix}$ when γ is fixed:

$$\begin{pmatrix} \mathbf{L} \\ \mathbf{b}^T \end{pmatrix} = \left(\frac{1}{n_s d_t} \begin{pmatrix} \mathbf{X}_s \\ \mathbf{1}^T \end{pmatrix} \begin{pmatrix} \mathbf{X}_s & \mathbf{1} \end{pmatrix} + \frac{\lambda_T}{d_s d_t} \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0}^T & 0 \end{pmatrix} \right)^{-1} \left(\frac{1}{n_s d_t} \begin{pmatrix} \mathbf{X}_s \\ \mathbf{1}^T \end{pmatrix} n_s \gamma \mathbf{X}_t + \frac{\lambda_T}{d_s d_t} \begin{pmatrix} \mathbf{I} \\ \mathbf{0}^T \end{pmatrix} \right). \tag{6}$$

Biased non-linear transformations In the biased non-linear case $\mathcal{H}$ becomes:

$$\mathcal{H} = \left\{ T : \exists\, \mathbf{L} \in \mathbb{R}^{n^s \times d^t}, \exists\, \mathbf{b} \in \mathbb{R}^{d^t}, \forall \mathbf{x}^s \in \Omega_{\mathcal{S}}, T(\mathbf{x}^s) = \begin{pmatrix} k_{\mathbf{X}_s}(\mathbf{x}^{sT}) & \mathbf{1} \end{pmatrix} \begin{pmatrix} \mathbf{L} \\ \mathbf{b}^T \end{pmatrix} \right\} \tag{7}$$

Optimization problem 1 can be rewritten as:

$$\underset{\begin{pmatrix} \mathbf{L} \\ \mathbf{b}^T \end{pmatrix} \in \mathbb{R}^{n^s+1 \times d^t}, \gamma \in \hat{\Pi}}{\arg\min} \frac{1}{n_s d_t} \left\| \begin{pmatrix} k_{\mathbf{X}_s}(\mathbf{X}_s) & \mathbf{1} \end{pmatrix} \begin{pmatrix} \mathbf{L} \\ \mathbf{b}^T \end{pmatrix} - n_s \gamma \mathbf{X}_t \right\|_{\mathcal{F}}^2 + \frac{\lambda_\gamma}{\max(\mathbf{C})} \langle \gamma, \mathbf{C} \rangle_{\mathcal{F}} + \frac{\lambda_T}{d_s d_t} \|k_{\mathbf{X}_s}(\cdot)\mathbf{L}\|_{\mathcal{F}}^2 . \tag{8}$$

As in the non biased case, it is possible to find a closed form solution for $\begin{pmatrix} \mathbf{L} \\ \mathbf{b}^T \end{pmatrix}$ when γ is fixed:

$$\begin{pmatrix} \mathbf{L} \\ \mathbf{b}^T \end{pmatrix} = \left(\frac{1}{n_s d_t} \begin{pmatrix} \mathbf{K}_{\mathbf{X}_s \mathbf{X}_s} \\ \mathbf{1}^T \end{pmatrix} \begin{pmatrix} \mathbf{K}_{\mathbf{X}_s \mathbf{X}_s} & \mathbf{1} \end{pmatrix} + \frac{\lambda_T}{d_s d_t} \begin{pmatrix} \mathbf{K}_{\mathbf{X}_s \mathbf{X}_s} & \mathbf{0} \\ \mathbf{0}^T & 0 \end{pmatrix} \right)^{-1} \frac{1}{n_s d_t} \begin{pmatrix} \mathbf{K}_{\mathbf{X}_s \mathbf{X}_s} \\ \mathbf{1}^T \end{pmatrix} n_s \gamma \mathbf{X}_t. \tag{9}$$

5 Proof of Equation (13) in the Main Paper

We recall the notations. Let T^* be the true transport map that we would obtain if we could solve Monge's problem. Let $\mathbf{B}_{\hat{\gamma}}$ be the empirical barycentric mapping of $\mathbf{X}_s$ using the probabilistic coupling $\hat{\gamma}$ learned between $\mathbf{X}_s$ and $\mathbf{X}_t$. Similarly let $\mathbf{B}_{\gamma_0}$ be the theoretical barycentric mapping associated with the probabilistic coupling γ_0 learned on $\mu_{\mathcal{S}}, \mu_{\mathcal{T}}$ the whole distributions and which corresponds to the solution of Kantorovich's problem. Using a slight abuse of notations we denote by $\mathbf{B}_{\hat{\gamma}}(\mathbf{x}^s)$ and $\mathbf{B}_{\gamma_0}(\mathbf{x}^s)$ the projection of $\mathbf{x}^s \in \mathbf{X}_s$ by these barycentric mappings. We have that:

$$\begin{aligned}
&\underset{\mathbf{x}^s \sim \Omega_{\mathcal{S}}}{\mathbb{E}} \|T(\mathbf{x}^s) - T^*(\mathbf{x}^s)\|_{\mathcal{F}}^2 \\
&\hfill \text{(Triangle inequality.)} \\
&\leq \underset{\mathbf{x}^s \sim \Omega_{\mathcal{S}}}{\mathbb{E}} \left(\|T(\mathbf{x}^s) - \mathbf{B}_{\gamma_0}(\mathbf{x}^s)\|_{\mathcal{F}} + \|\mathbf{B}_{\gamma_0}(\mathbf{x}^s) - T^*(\mathbf{x}^s)\|_{\mathcal{F}} \right)^2 \\
&\hfill ((a+b)^2 \leq 2a^2 + 2b^2.) \\
&\leq 2 \underset{\mathbf{x}^s \sim \Omega_{\mathcal{S}}}{\mathbb{E}} \|T(\mathbf{x}^s) - \mathbf{B}_{\gamma_0}(\mathbf{x}^s)\|_{\mathcal{F}}^2 + 2 \underset{\mathbf{x}^s \sim \Omega_{\mathcal{S}}}{\mathbb{E}} \|\mathbf{B}_{\gamma_0}(\mathbf{x}^s) - T^*(\mathbf{x}^s)\|_{\mathcal{F}}^2
\end{aligned}$$

Furthermore considering that $\mathcal{H}$ is as proposed in the paper and using Theorem 2 in [4] we have with high probability that:

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\mathbf{x}^s\sim\Omega_{\mathcal{S}}} \|T(\mathbf{x}^s) - T^*(\mathbf{x}^s)\|_{\mathcal{F}}^2 \\
&\qquad \leq 2 \sum_{\mathbf{x}^s\in\mathbf{X}_s} \|T(\mathbf{x}^s) - \mathbf{B}_{\gamma_0}(\mathbf{x}^s)\|_{\mathcal{F}}^2 + \mathcal{O}\left(\frac{1}{\sqrt{n_s}}\right) + 2 \mathop{\mathbb{E}}_{\mathbf{x}^s\sim\Omega_{\mathcal{S}}} \|\mathbf{B}_{\gamma_0}(\mathbf{x}^s) - T^*(\mathbf{x}^s)\|_{\mathcal{F}}^2 \\
&\qquad\qquad \text{(Triangle inequality.)} \\
&\qquad \leq 2 \sum_{\mathbf{x}^s\in\mathbf{X}_s} \left(\|T(\mathbf{x}^s) - \mathbf{B}_{\hat{\gamma}}(\mathbf{x}^s)\|_{\mathcal{F}} + \|\mathbf{B}_{\hat{\gamma}}(\mathbf{x}^s) - \mathbf{B}_{\gamma_0}(\mathbf{x}^s)\|_{\mathcal{F}}\right)^2 + \mathcal{O}\left(\frac{1}{\sqrt{n_s}}\right) \\
&\qquad\quad + 2 \mathop{\mathbb{E}}_{\mathbf{x}^s\sim\Omega_{\mathcal{S}}} \|\mathbf{B}_{\gamma_0}(\mathbf{x}^s) - T^*(\mathbf{x}^s)\|_{\mathcal{F}}^2 \\
&\qquad\qquad ((a+b)^2 \leq 2a^2 + 2b^2.) \\
&\qquad \leq 4 \sum_{\mathbf{x}^s\in\mathbf{X}_s} \|T(\mathbf{x}^s) - \mathbf{B}_{\hat{\gamma}}(\mathbf{x}^s)\|_{\mathcal{F}}^2 + \mathcal{O}\left(\frac{1}{\sqrt{n_s}}\right) \\
&\qquad\quad + 4 \sum_{\mathbf{x}^s\in\mathbf{X}_s} \|\mathbf{B}_{\hat{\gamma}}(\mathbf{x}^s) - \mathbf{B}_{\gamma_0}(\mathbf{x}^s)\|_{\mathcal{F}}^2 + 2 \mathop{\mathbb{E}}_{\mathbf{x}^s\sim\Omega_{\mathcal{S}}} \|\mathbf{B}_{\gamma_0}(\mathbf{x}^s) - T^*(\mathbf{x}^s)\|_{\mathcal{F}}^2 . \qquad (10)
\end{aligned}$$

6 Complementary Information on Experimental Protocol for the Domain Adaptation Experiments

Algorithm 2 explains the 2-fold circular validation used for tuning the hyper-parameters and inspired from [5, 6]. In this algorithm M is any model able to bring closer the source and the target. For example, with our linear mapping learned from our regularized OT formulation, we have $M(\mathbf{X}_t) = \mathbf{X}_t$ and $M(\mathbf{X}_s) = \mathbf{X}_s\mathbf{L}$.

Algorithm 2: Circular validation.

```
input : $(\mathbf{X}_s, \mathbf{y}_s)$ source examples and their labels, $\mathbf{X}_t$ target examples, $\mathcal{A}_{\boldsymbol{\lambda}}$ a learning procedure using
        hyper-parameters $\boldsymbol{\lambda}$.
output: Average accuracy of $\mathcal{A}_{\boldsymbol{\lambda}}$.
begin
    Split $(\mathbf{X}_s, \mathbf{y}_s)$ in two halves $(\mathbf{X}_s^1, \mathbf{y}_s^1)$ and $(\mathbf{X}_s^2, \mathbf{y}_s^2)$.
    Learn $M^1 = \mathcal{A}_{\boldsymbol{\lambda}}(\mathbf{X}_s^1, \mathbf{y}_s^1, \mathbf{X}_t)$ and set $\mathbf{y}_t^1$ the pseudo-labels of $M^1(\mathbf{X}_t)$ obtained from a 1NN
    learned on $(M^1(\mathbf{X}_s^1), \mathbf{y}_s^1)$.
    Set $s^1$ the accuracy of a 1NN learned on $(M^1(\mathbf{X}_t), \mathbf{y}_t^1)$ and evaluated on $(M^1(\mathbf{X}_s^2), \mathbf{y}_s^2)$ .
    Learn $M^2 = \mathcal{A}_{\boldsymbol{\lambda}}(\mathbf{X}_s^2, \mathbf{y}_s^2, \mathbf{X}_t)$ and set $\mathbf{y}_t^2$ the pseudo-labels of $M^2(\mathbf{X}_t)$ obtained from a 1NN
    learned on $(M^2(\mathbf{X}_s^2), \mathbf{y}_s^2)$.
    Set $s^2$ the accuracy of a 1NN learned on $(M^2(\mathbf{X}_t), \mathbf{y}_t^2)$ and evaluated on $(M^2(\mathbf{X}_s^1), \mathbf{y}_s^1)$ .
    return $\frac{s^1+s^2}{2}$.
```

7 Illustrations on the Moons Dataset for the Domain Adaptation Experiments

In Figure 1 we propose some illustrations of the transformation learned by our approach on the Moons dataset.

8 Complementary Information on Gradient Adaptation in Image Editing

In the paper, we build from a technique, denoted *Poisson Image Editing*, that operates in the gradient domain of the image. Hence, the gradients of the selection operate as a guidance field for an image reconstruction based on membrane interpolation with appropriate boundary conditions extracted from

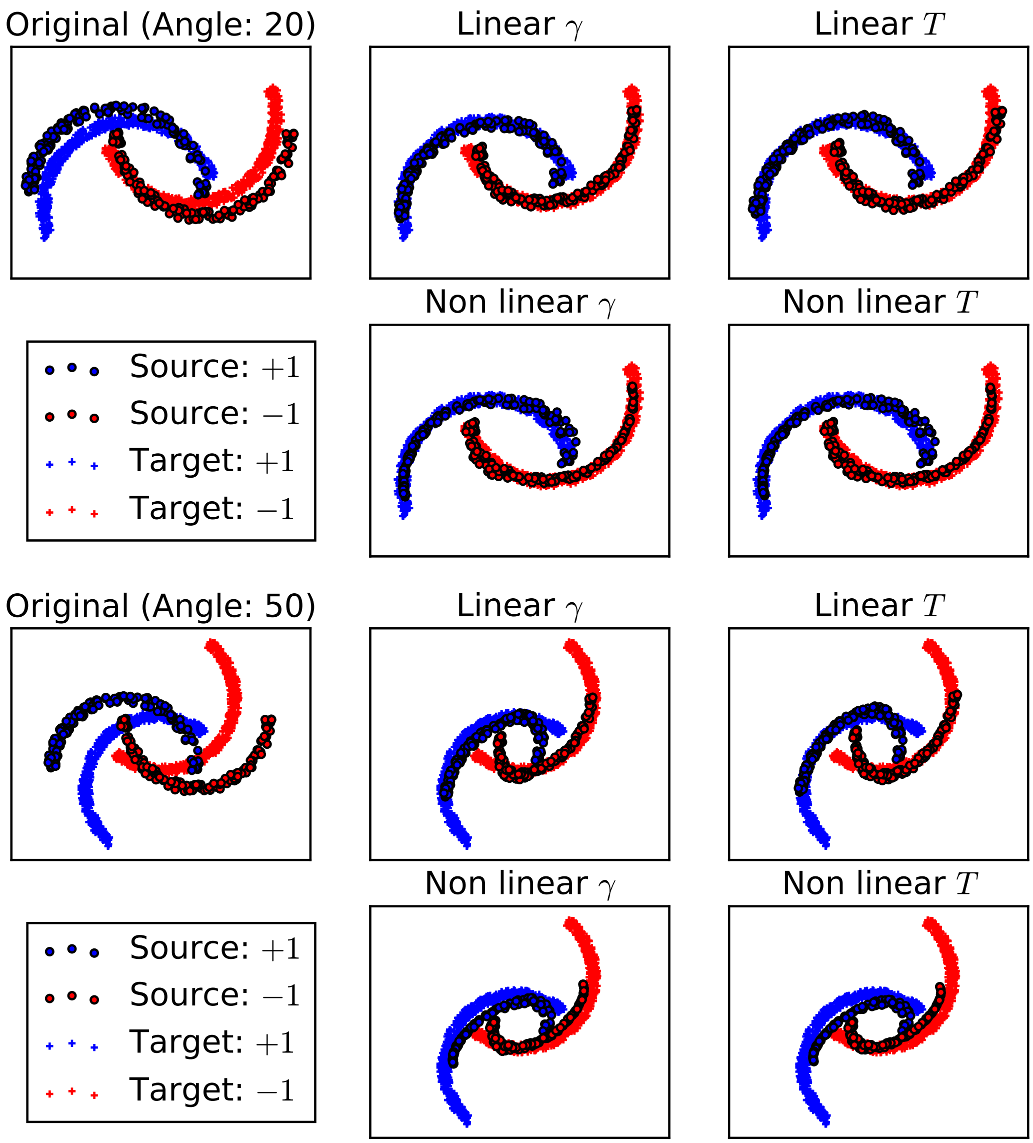

Figure 1: Illustrations of our approach on the Moons dataset when the rotation is of 20 degrees (first and second rows) and 50 degrees (third and fourth rows). The transformation T follows closely the transport map γ and the shapes of the two moons are well preserved. Furthermore learning a linear transformation is better when the angle is 50 degrees. The shrinkage effect is due to the regularization on the transformation which penalizes complex solutions.

the target image. Let f be an unknown scalar function (usually a component of the color space of the image) defined on a given region of the image Ω. Let f_t be the target image defined everywhere apart from the interior of Ω. The Poisson editing method operates by solving for f the following variational optimization problem with Dirichlet boundary conditions:

$$\min_f \int\int_{\Omega} |\nabla f - \mathbf{v}|^2 \quad \text{with} \quad f|_{\partial\Omega} = f_t|_{\partial\Omega}. \tag{11}$$

Here, $\mathbf{v}$ is the guidance field, which is usually given as the gradient from the source image f_s over the domain Ω, *i.e.* $\mathbf{v} = \nabla f_s|_{\Omega}$. One can show that the unique solution to this problem is the solution of the following Poisson equation [7]:

$$\Delta f = \text{div } \mathbf{v} \quad \text{over } \Omega, \quad \text{with} \quad f|_{\partial\Omega} = f_t|_{\partial\Omega}, \tag{12}$$

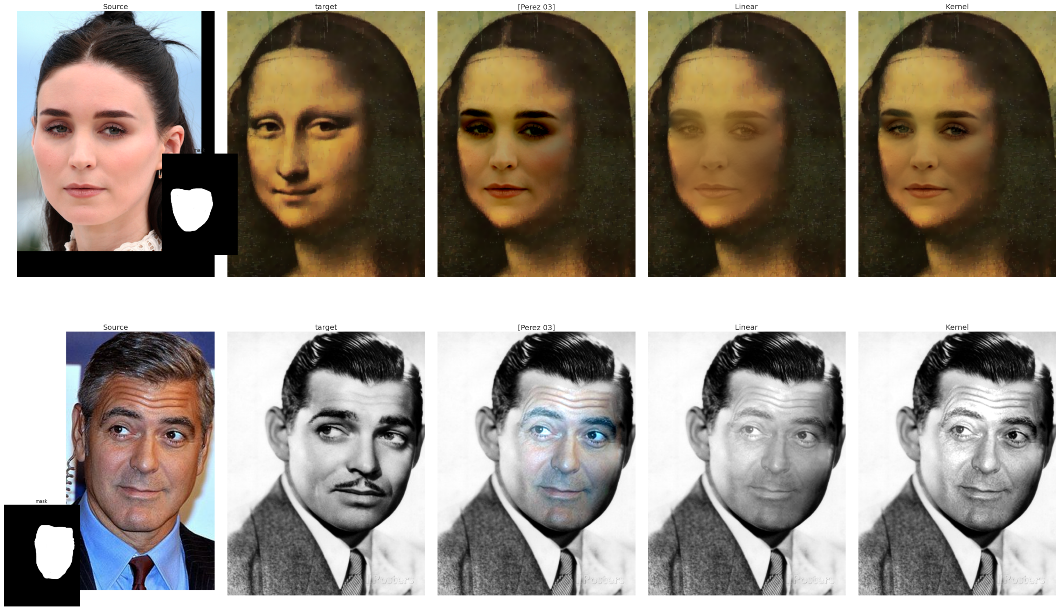

Figure 2: Complementary Illustrations of seamless copies with gradient adaptation.

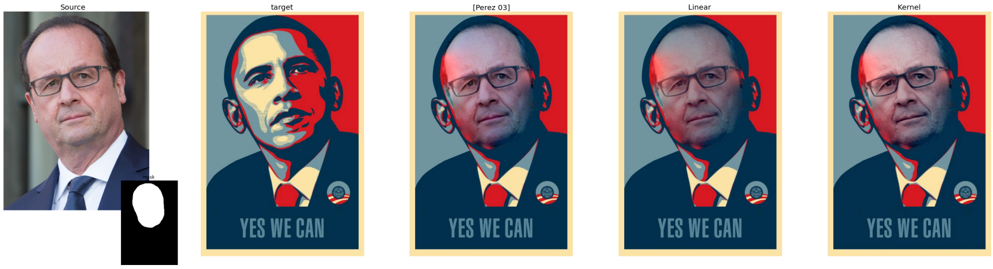

Figure 3: Illustration of failure of style adaptation.

where div stands for the divergence operator. Using appropriate first order discretization of the Laplacian operator, solving for this problem amounts to solve a big sparse linear system, which can be performed efficiently with multi-grid solvers. We propose in the paper to enhance the generality of this technique by forcing the gradient distribution from the source image to follow the gradient distribution in the target image. We start by learning a transfer function $T_{s\to t} : \mathbb{R}^6 \to \mathbb{R}^6$. We then solve for the following system:

$$\Delta f = \text{div}\ T_{s\to t}(\mathbf{v}) \quad \text{over } \Omega, \quad \text{with} \quad f|_{\partial\Omega} = f_t|_{\partial\Omega}. \tag{13}$$

In Figure 2 and 3 we show other results produced by our methods. Figure 3 illustrates one particular case of failure of style adaptation: as our method does not modify the spatial arrangement of the gradient, it is not possible to produce the same vast swaths of colors as in the target image.

References

[1] S. Ferradans, N. Papadakis, G. Peyré, and J.-F. Aujol. Regularized discrete optimal transport. *SIIMS*, 2014.

[2] M. Frank and P. Wolfe. An algorithm for quadratic programming. *NRL*, 3(1-2), 1956.

[3] M. Jaggi. Revisiting Frank-Wolfe: Projection-free sparse convex optimization. In *ICML*, 2013.

[4] M. Perrot and A. Habrard. Regressive virtual metric learning. In *NIPS*, 2015.

[5] L. Bruzzone and M. Marconcini. Domain adaptation problems: A DASVM classification technique and a circular validation strategy. *IEEE PAMI*, 32(5), 2010.

[6] E. Zhong, W. Fan, Q. Yang, O. Verscheure, and J. Ren. Cross validation framework to choose amongst models and datasets for transfer learning. In *ECML PKDD*, 2010.

[7] P. Pérez, M. Gangnet, and A. Blake. Poisson image editing. *ACM Trans. on Graphics*, 22(3), 2003.