SIAM J. SCI. COMPUT.
Vol. 39, No. 4, pp. A1793–A1810

PRECONDITIONING OF OPTIMAL TRANSPORT*

MAX KUANG† AND ESTEBAN G. TABAK†

Abstract. A preconditioning procedure is developed for the L_2 and more general optimal transport problems. The procedure is based on a family of affine map pairs which transforms the original measures into two new measures that are closer to each other while preserving the optimality of solutions. It is proved that the preconditioning procedure minimizes the remaining transportation cost among all admissible affine maps. The procedure can be used on both continuous measures and finite sample sets from distributions. In numerical examples, the procedure is applied to multivariate normal distributions, to a two-dimensional shape transform problem, and to color-transfer problems.

Key words. preconditioning, optimal transport, matrix factorization

AMS subject classifications. 65F08, 15A23, 49K45

DOI. 10.1137/16M1074953

1. Introduction.

The original optimal transport problem, proposed by Monge in 1781 [17], addresses the displacement of a pile of soil between two locations with minimal cost. Giving the cost $c(x, y)$ of moving a unit mass from point x to point y, one seeks the map $y = T(x)$ that minimizes its integral. After normalizing the two piles so that each has total mass one, so that they can be regarded as probability measures, the problem adopts the form

$$\min_{T_\sharp \mu = \nu} \int c(x, T(x)) d\mu(x), \tag{1}$$

where μ and ν are the original and target measures, and $T_\sharp \mu$ denotes the pushforward measure of μ by the map T.

In the 20th century, Kantorovich [13] relaxed Monge's problem, allowing the movement of soil from one to multiple locations and vice versa. Denoting the mass transported from x to y by $\pi(x, y)$, the minimization problem can be rewritten as

$$\min_{\pi} \int c(x, y)\pi(x, y) dx dy \tag{2}$$

among couplings $\pi(x, y)$ satisfying the marginal constraints

$$\int \pi(x, y) dy = \mu(x),$$
$$\int \pi(x, y) dx = \nu(y).$$

Since the second half of the 20th century, mathematical properties of the optimal transport solution have been studied extensively, as well as applications in many different areas (see, for instance, [20, 15, 4, 8, 11, 5], or [26] for a comprehensive list).

*Submitted to the journal's Methods and Algorithms for Scientific Computing section May 12, 2016; accepted for publication (in revised form) May 17, 2017; published electronically August 31, 2017.

http://www.siam.org/journals/sisc/39-4/M107495.html

Funding: This work was supported by grants from the Office of Naval Research and the NYU-AIG Partnership on Global Resilience.

†Courant Institute, New York University, New York, NY 10012 (kuang@cims.nyu.edu, tabak@cims.nyu.edu).

Since closed-form solutions of the multidimensional optimal transport problems are relatively rare, a number of numerical algorithms have been proposed. Some recent representative examples of the different approaches follow.

PDE methods: Benamou and Brenier [3] introduced a computational fluid approach to solve the problem with continuous distributions $\mu_{1,2}$, exploiting the structure of the interpolant of the optimal map to solve the PDE corresponding to the optimization problem in the dual variables.

Adaptive linear programming: Oberman and Ruan [18] discretized the given continuous distributions and solved the resulting linear programming problem in an adaptive way that exploits the sparse nature of the solution (the fact that the optimal plan has support on a map).

Entropy regularization: The discrete version of optimal transport is the earth mover's problem in image processing [22], a linear programming problem widely used to measure the distance between images and in networks. Recent development on entropy regularization [23] introduced effective algorithms to solve regularized versions of these problems.

Data-driven formulations: Data-driven formulations take as input not the distributions $\mu_{1,2}$ but sample sets from both. Methodologies proposed include a fluid-flow-like algorithm [25], an adaptive linear programming approach [6], and a procedure based on approximating the interpolant in a feature-space [14].

In this paper, we introduce a procedure to precondition the input probability measures or samples thereof, so that the resulting measures or sample sets are closer to each other while preserving the optimality of maps. More precisely, affine maps F and G are found that transform the original measures μ, ν into two new ones $\tilde{\mu}, \tilde{\nu}$ with identical mean and covariance, with the key additional property that the transformation $Y = G^{-1}(T(F(X)))$, which composes the preconditioner and the optimal map T between $\tilde{\mu}$ and $\tilde{\nu}$, is the optimal map between the two original distributions μ and ν. The procedure and its properties are discussed for the standard square-distance cost and for the more general class of cost functions induced by an inner product.

In theoretical applications, the preconditioning procedure is used to give alternative derivations of a lower bound for the total transportation cost and of the optimal map between multivariate normal distributions. For practical applications, we use the procedure on sample sets to get preconditioned sets, which are then given as input to optimal transport algorithms to calculate the optimal map. Then, inverting the the preconditioning map pairs used, we recover the optimal map between the original distributions.

Through analysis and numerical examples, we see that the preconditioning procedure can often explain and remove a substantial portion of the transportation cost with little computational cost. Furthermore, assuming that the following numerical optimal transport algorithm finds problems with closer measures easier to solve, the preconditioning algorithm will also increase the performance of the underlying algorithm.

2. Optimal transport.

We first let μ and ν be two probability measures on the same sample space $\mathcal{X}$ and let X be a random variable from μ. Optimal transport asks how to optimally move the mass from μ to ν, given a function $c(x, y)$ that represents the cost of moving a unit of mass from point x to point y. Monge's formulation seeks a map $y = T(x)$ that minimizes the total transportation cost:

$$\min_{X \sim \mu, T_\sharp \mu = \nu} \mathbb{E}\ c(X, T(X)), \tag{3}$$

where $T_\sharp \mu$ represents the pushforward measure of μ through the map T.

A transfer plan $\pi(x,y)$ is the law of a coupling (X,Y) between the two measures μ and ν. For any measurable set $E \subset \mathcal{X}$,

$$\pi(E \times \mathcal{X}) = \mu(E), \quad \pi(\mathcal{X} \times E) = \nu(E).$$

Denoting the family of all transfer plans by $\Pi(\mu,\nu)$, Kantorovich's relaxation of the optimal transport problem is

$$\min_{\pi \in \Pi(\mu,\nu)} \mathbb{E}\; c(X,Y). \tag{4}$$

Since the maps $Y = T(X)$ represent a subset of all couplings between μ and ν, the feasible domain for (3) lies within the one for (4).

While there are many results on the general optimal transport problem, a particularly well studied and useful case is the L_2 optimal transport on $\mathcal{R}^N$, in which μ and ν are probability measures on $\mathcal{R}^N$ and the cost function $c(x,y)$ is given by the square Euclidean distance $\|x-y\|^2$. In this case, with moderate requirements, one can prove that the solution to Kantorovich's relaxation (4) is unique and agrees with the solution to Monge's problem (3). In other words, the unique optimal coupling (X,Y) corresponds to a map $Y = T(X)$. Moreover, this optimal map is the gradient of a convex potential ϕ. These properties have been introduced in numerous works, for example, in Theorem 3.1 in [9].

THEOREM 2.1. *For Kantorovich's relaxation* (4) *with the quadratic cost function and absolute continuous measures μ and ν, the optimal coupling (X,Y) is a map $Y = T(X)$, where $T : \mathcal{R}^N \to \mathcal{R}$ is defined by*

$$T(x) = \nabla\phi(x), \tag{5}$$

where $\phi(x)$ is convex and $T_\sharp\mu = \nu$.

While this characterization of the solution is attractively simple, closed-form solutions of the L_2 optimal transport on $\mathcal{R}^N$ are rare for $N > 1$. The difficulties of deriving closed-form solutions prompted research to solve the optimal transport problem numerically. An incomplete list of formulations and methods can be found in section 1.

The goal of this paper is not to provide a complete numerical recipe to solve the L_2 optimal transport problem, but to introduce a practical preconditioning procedure. This procedure transforms the original measures μ and ν into two new measures, so that the optimal transport problem between the new measures is easier to solve, while the optimality of solutions is preserved by the transformation. The procedure extends beyond the square-distance cost to all cost functions induced by an inner product.

3. Admissible map pairs. The basic framework of the preconditioning procedure is as follows:

$$\begin{array}{ccc}
X \sim \mu & \xrightarrow{Y=G^{-1}(T(F(X)))} & Y \sim \nu \\
\Big\downarrow {\scriptstyle \tilde{X}=F(X)} & & \Big\downarrow {\scriptstyle \tilde{Y}=G(Y)} \\
\tilde{X} \sim \tilde{\mu} & \xrightarrow{\tilde{Y}=T(\tilde{X})} & \tilde{Y} \sim \tilde{\nu}.
\end{array}$$

Suppose that we transform μ and ν into two new measures $\tilde{\mu}$ and $\tilde{\nu}$ via some invertible maps F and G and that the solution to the new L_2 optimal transport problem between $\tilde{\mu}$ and $\tilde{\nu}$ is given by $\tilde{Y} = T(\tilde{X})$. Then the map

$$Y = G^{-1}(T(F(X))) \tag{6}$$

pushes forward μ into ν, yet it is generally not optimal. We call the pair of invertible maps (F, G) an *admissible map pair* if the resulting map (6) is always optimal for the original problem between μ and ν.

There are several simple *admissible map pairs.*

DEFINITION 3.1 (translation pairs). *Given two vectors m_1, m_2 in $\mathcal{R}^N$, a translation pair (F, G) is defined by*

$$F(X) = X - m_1, \quad G(Y) = Y - m_2. \tag{7}$$

If $\tilde{Y} = T(\tilde{X})$ is an optimal map, then $T = \nabla\phi$ for some convex function ϕ, which implies that

$$Y = m_2 + T(X - m_1) = \nabla\left[m_2 X + \phi(X - m_1)\right], \tag{8}$$

so $Y = G^{-1}(T(F(X)))$ is indeed the optimal map between μ and ν. Thus *translation pairs* are *admissible map pairs.*

Among all *translation pairs*, we can minimize the total transportation cost in the transformed problem:

$$\begin{aligned}
\mathbb{E}\|\tilde{X} - \tilde{Y}\|^2 &= \mathbb{E}\|X - m_1 - Y + m_2\|^2 \\
&= \mathbb{E}\|X - \mathbb{E}X - Y + \mathbb{E}Y\|^2 + \|\mathbb{E}X - m_1 - \mathbb{E}Y + m_2\|^2 \\
&\geq \mathbb{E}\|X - \mathbb{E}X - Y + \mathbb{E}Y\|^2.
\end{aligned}$$

This shows that the transportation cost between $\tilde{X}$ and $\tilde{Y}$ is minimized when $\mathbb{E}X - \mathbb{E}Y = m_1 - m_2$. In particular, we can adopt $m_1 = \mathbb{E}X$ and $m_2 = \mathbb{E}Y$, which gives both measures a zero mean. We call the corresponding *translation pair* the *mean translation pair.*

DEFINITION 3.2 (scaling pairs). *Given two nonzero numbers α, β in $\mathcal{R}$, the scaling pair (F, G) is defined by*

$$F(X) = \alpha X, \quad G(Y) = \beta Y. \tag{9}$$

Clearly if $\tilde{Y} = T(\tilde{X}) = \nabla\phi(\tilde{X})$ is an optimal map, then

$$Y = \frac{1}{\beta}T(\alpha X) = \nabla\frac{1}{\alpha\beta}\phi(\alpha X) \tag{10}$$

is also an optimal map. So all the *scaling pairs* are *admissible map pairs.* In particular, if X and Y are not constants, one can choose

$$\alpha = \frac{1}{\sqrt{\mathbb{E}\|X\|^2}}, \quad \beta = \frac{1}{\sqrt{\mathbb{E}\|Y\|^2}},$$

so that

$$\mathbb{E}\|\tilde{X}\|^2 = \mathbb{E}\|\tilde{Y}\|^2 = 1.$$

We call this specific *scaling pair* the *normalizing scaling pair.*

Next we discuss general linear *admissible map pairs.* We will think of X as row vectors, so the matrices representing linear transformations act on X on the right.

THEOREM 3.3. *Let $F(X) = XA$ and $G(Y) = YB$, where $A, B \in \mathcal{R}^{N\times N}$ are invertible matrices. Denote by $\tilde{Y} = T(\tilde{X})$ the optimal map from $\tilde{\mu}$ to $\tilde{\nu}$. If $B = (A^T)^{-1}$, the induced map between μ and ν is also optimal.*

Proof. The induced map can be written as

$$Y = T(XA)B^{-1} = T(XA)A^T.$$

Let $T(X) = \nabla\phi(X)$ and $\psi(X) = \phi(XA)$; then we have

$$Y_i = \sum_{j=1}^{N} \nabla\phi(XA)_j (A^T)_{ij} = \frac{\partial}{\partial X_i}\phi(XA) \ \Rightarrow\ Y = \nabla\psi(X). \tag{11}$$

Since ψ is also a convex function, the induced map $Y = T(XA)B^{-1}$ is also an optimal map. ∎

Remark 3.4. Another way to understand this theorem is to consider map pairs (F, G) that do not alter the inner product. In fact, the theorem holds if for any $x, y \in \mathcal{R}^N$,

$$xy^T = F(x)G(y)^T. \tag{12}$$

This observation implies that the same result holds for more general cost functions: as long as the metric $d(x, y)$ is induced by an inner product $\langle x, y\rangle$, we only need the pair F and G to be adjoint operators to guarantee they form an *admissible map pair*.

The above theorem gives us a family of new *admissible map pairs*.

DEFINITION 3.5 (linear pairs). *Let A be an invertible matrix in $\mathcal{R}^{N\times N}$; then the linear pair (F, G) is defined by*

$$F(X) = XA, \quad G(Y) = Y(A^T)^{-1}. \tag{13}$$

We first give some examples of common *linear pairs*.

DEFINITION 3.6 (orthogonal pairs). *For any orthogonal matrix A,*

$$F(X) = XA, \quad G(Y) = YA \tag{14}$$

is called a orthogonal map pair.

For orthogonal pairs, we have $(A^T)^{-1} = A$. This means that performing the same orthogonal linear transformation on both measures preserves the optimality of solutions. The interpretation of this result is straightforward, as an orthogonal map yields a distance-preserving coordinate change which does not alter the cost function.

DEFINITION 3.7 (stretching pairs). *For any unit vector d and scalar α, we can stretch X by a factor of α along d, and at the same time stretch Y by a factor of $1/\alpha$ along the same direction:*

$$\begin{aligned} F(X) &= X - (Xd^T)d + \alpha(Xd^T)d = X(I + (\alpha - 1)d^T d),\\ G(Y) &= Y - (Yd^T)d + 1/\alpha(Xd^T)d = X(I + (1/\alpha - 1)d^T d). \end{aligned} \tag{15}$$

We call such map pairs stretching pairs.

It can be verified this is indeed a *linear pair*, and thus an *admissible map pair*.

Composing translation and linear pairs, one obtains a more general class of *affine pairs*. Among all *affine pairs*, we seek the optimal one for our preconditioning procedure. We first state a linear algebra result.

THEOREM 3.8. *For any two positive definite matrices Σ_1 and Σ_2 in $R^{N\times N}$, there exists an invertible matrix $A \in R^{N\times N}$ such that*

$$D = A^T \Sigma_1 A = A^{-1}\Sigma_2 (A^T)^{-1}, \tag{16}$$

where D is a diagonal matrix with entries satisfying

$$d_1 \geq d_2 \geq \cdots \geq d_N > 0. \tag{17}$$

In addition, D is unique.

Proof. We first prove the existence of A. Since $\Sigma_1^{1/2}$ is invertible, we can replace A by a matrix B satisfying

$$B = \Sigma_1^{1/2} A$$

and

$$D = B^T B = B^{-1}\Sigma_1^{1/2}\Sigma_2\Sigma_1^{1/2}(B^T)^{-1}.$$

Because $\Sigma_1^{1/2}\Sigma_2\Sigma_1^{1/2}$ is positive definite, it admits an eigenvalue decomposition of the form

$$\Sigma_1^{1/2}\Sigma_2\Sigma_1^{1/2} = Q\Lambda Q^T, \tag{18}$$

with Q orthogonal and Λ diagonal with sorted, positive diagonal entries. Setting $B = Q\Lambda^{1/4}$, we have

$$B^T B = \Lambda^{1/2}$$

and

$$B^{-1}\Sigma_1^{1/2}\Sigma_2\Sigma_1^{1/2}(B^T)^{-1} = \Lambda^{-1/4}Q^T Q\Lambda Q^T Q\Lambda^{-1/4} = \Lambda^{1/2}.$$

Thus the conditions of the theorem are satisfied with

$$D = \Lambda^{1/2}, \quad A = \Sigma_1^{-1/2} Q\Lambda^{1/4}. \tag{19}$$

To prove the uniqueness of D, suppose that there are D_1, A_1 and D_2, A_2 such that

$$D_1 = A_1^T\Sigma_1 A_1 = A_1^{-1}\Sigma_2(A_1^T)^{-1},$$
$$D_2 = A_2^T\Sigma_1 A_2 = A_2^{-1}\Sigma_2(A_2^T)^{-1}.$$

Then

$$D_1^2 = A_1^{-1}\Sigma_2\Sigma_1 A_1,$$
$$D_2^2 = A_2^{-1}\Sigma_2\Sigma_1 A_2,$$

implying that D_1^2 , $\Sigma_2\Sigma_1$, and D_2^2 are similar to one other. Since D_1 and D_2 are positive diagonal matrices with sorted entries, they must be identical, which proves the uniqueness of D. ∎

Using the theorem above, we can define the following *optimal linear pair*.

DEFINITION 3.9 (optimal linear pair). *Assume that μ and ν are mean-zero measures with covariance matrices Σ_1 and Σ_2, and let A be an $N \times N$ matrix that satisfies* (16)*. We define the optimal linear pair (F, G) by*

$$F(X) = XA, \quad G(Y) = Y(A^T)^{-1}. \tag{20}$$

(Notice that the matrix A can be constructed following (18) *and* (19) *in the proof of Theorem* 3.8.*)*

This pair has the following useful properties.

PROPERTY 3.10. *The resulting random variables $\tilde{X}$, $\tilde{Y}$ derived from the optimal linear pair have the same diagonal covariance matrix D:*

$$\mathbb{E}\tilde{X}^T\tilde{X} = A^T\Sigma_1 A = D, \tag{21}$$

$$\mathbb{E}\tilde{Y}^T\tilde{Y} = A^{-1}\Sigma_2(A^T)^{-1} = D. \tag{22}$$

PROPERTY 3.11. *Among all possible linear pairs $X' = XC$, $Y' = Y(C^T)^{-1}$ given by an invertible matrix C, the optimal linear pair minimizes $\mathbb{E}\|X' - Y'\|^2$. In other words, for any invertible matrix C,*

$$\mathbb{E}\|X' - Y'\|^2 \geq \mathbb{E}\|\tilde{X} - \tilde{Y}\|^2. \tag{23}$$

Proof. For any matrix C, we have

$$\begin{aligned}
\mathbb{E}\|X' - Y'\|^2 &= \mathbb{E}X'X'^T + \mathbb{E}Y'Y'^T - 2\mathbb{E}X'Y'^T \\
&= \mathbb{E}XCC^TX^T + \mathbb{E}Y(C^T)^{-1}C^{-1}Y^T - 2\mathbb{E}XY^T \\
&= \mathbb{E}\,\mathrm{tr}(C^TX^TXC) + \mathbb{E}\,\mathrm{tr}(C^{-1}Y^TY(C^T)^{-1}) - 2\mathbb{E}XY^T \\
&= \mathrm{tr}(C^T\Sigma_1 C) + \mathrm{tr}(C^{-1}\Sigma_2(C^T)^{-1}) - 2\mathbb{E}XY^T.
\end{aligned}$$

On the other hand, (16) is equivalent to

$$\Sigma_1 = (A^T)^{-1}DA^{-1}, \quad \Sigma_2 = ADA^T.$$

In terms of $S = A^{-1}C$,

$$\begin{aligned}
\mathbb{E}\|X' - Y'\|^2 &= \mathrm{tr}(S^TDS) + \mathrm{tr}(S^{-1}D(S^T)^{-1}) - 2\mathbb{E}XY^T \\
&= \mathrm{tr}(SS^TD) + \mathrm{tr}((SS^T)^{-1}D) - 2\mathbb{E}XY^T.
\end{aligned}$$

Writing $S = (s_1, s_2, \ldots, s_N)^T$ and $(S^T)^{-1} = (z_1, z_2, \ldots, z_N)^T$, we have

$$\begin{aligned}
\mathbb{E}\|X' - Y'\|^2 &= \sum_{i=1}^N d_i s_i^T s_i + \sum_{i=1}^N d_i z_i^T z_i - 2\mathbb{E}XY^T \\
&= \sum_{i=1}^N d_i(s_i^T s_i + z_i^T z_i) - 2\mathbb{E}XY^T \\
&\geq \sum_{i=1}^N d_i(2s_i^T z_i) - 2\mathbb{E}XY^T \\
&= 2\sum_{i=1}^N d_i - 2\mathbb{E}XY^T \\
&= \mathbb{E}\|\tilde{X} - \tilde{Y}\|^2.
\end{aligned}$$

Notice that we have the equal sign when $S = I$, which means that $C = A$. Thus

$$\mathbb{E}\|X' - Y'\|^2 \geq 2\sum_{i=1}^{N} d_i - 2\mathbb{E}XY^T = \mathbb{E}\|\tilde{X} - \tilde{Y}\|^2. \qquad \square$$

Composing the *mean translation pair* and the *optimal linear pair*, one obtains the *optimal affine pair*. It follows from the properties above that the *optimal affine pair* not only gives the two distributions zero means and transforms the covariance matrices into diagonal matrices, but also minimizes the distance between $\tilde{\mu}$ and $\tilde{\nu}$ among all *affine pairs*.

4. General quadratic cost functions. In Theorem 3.3, we introduced a class of affine maps that preserves the optimality of solutions for the square-distance cost. As mentioned in Remark 3.4, similar results hold for the more general cost functions induced by an inner product. We have the following generalization of Theorem 3.3.

THEOREM 4.1. *Let $\langle \cdot, \cdot \rangle$ be an inner product in $\mathcal{X}$. For the optimal transport problem with cost*

$$c(x, y) = \langle x - y, x - y \rangle, \tag{24}$$

we have that (F, G) is an admissible map pair if F and G are adjoint operators with respect to inner product $\langle \cdot, \cdot \rangle$.

Proof. It follows from the fact that $c(x, y) = \|x\|^2 + \|y\|^2 - 2\langle x, y \rangle$, where only the last term depends on the actual coupling between X and Y, that

$$\operatorname{argmin} \mathbb{E}\left[c(X, Y)\right] = \operatorname{argmax} \mathbb{E}\left[\langle X, Y \rangle\right].$$

Since this applies to both the original and the preconditioned problems, their optimal solutions satisfy

$$(X^*, Y^*) = \operatorname{argmax}\left[\mathbb{E}\langle X, Y \rangle\right] \quad \text{and} \quad (\tilde{X}^*, \tilde{Y}^*) = \operatorname{argmax}[\mathbb{E}\langle \tilde{X}, \tilde{Y} \rangle].$$

But if F and G are adjoint,

$$\langle \tilde{X}, \tilde{Y} \rangle = \langle F(X), G(Y) \rangle = \langle X, Y \rangle,$$

so

$$(\tilde{X}^*, \tilde{Y}^*) = (F(X^*), G(Y^*)),$$

proving the conclusion. $\square$

Any inner product on $\mathcal{R}^N$ can be written in terms of the standard vector multiplication, through the introduction of a positive definite kernel matrix K:

$$\langle x, y \rangle = xKy^T, \tag{25}$$

so stating that the linear operators $F(X) = XA$, $G(Y) = YB$ are adjoint is equivalent to

$$AKB^T = K. \tag{26}$$

We can also derive the *optimal linear pair* for general cost functions. Here we only state without proof the core linear algebra theorem.

THEOREM 4.2. *Let Σ_1, Σ_2, and K be positive definite matrices in $R^{N\times N}$. There exist invertible matrices $A, B \in R^{N\times N}$ such that*

$$AKB^T = K \tag{27}$$

and

$$D = K^{1/2}A^T\Sigma_1 AK^{1/2} = K^{1/2}B^T\Sigma_2 BK^{1/2}, \tag{28}$$

where D is a unique diagonal matrix with entries satisfying

$$d_1 \geq d_2 \geq \cdots \geq d_N > 0. \tag{29}$$

Matrices constructed so as to satisfy the above theorem give the *optimal linear pairs* with respect to the corresponding cost. Notice that in this case the resulting measures no longer have diagonal covariance matrices:

$$\mathbb{E}\tilde{X}^T\tilde{X} = \mathbb{E}\tilde{Y}^T\tilde{Y} = K^{-1/2}DK^{-1/2}. \tag{30}$$

5. Preconditioning procedure and its applications. Returning to the standard cost, we introduce the full preconditioning procedure using all the *admissible map pairs* discussed in section 3.

DEFINITION 5.1 (preconditioning procedure). *For two random variables X and Y with probability measures μ and ν, let*

$$m_1 = \mathbb{E}X, \quad m_2 = \mathbb{E}Y, \tag{31}$$

$$\Sigma_1 = \mathbb{E}\left[(X-m_1)^T(X-m_1)\right], \quad \Sigma_2 = \mathbb{E}\left[(Y-m_2)^T(Y-m_2)\right]. \tag{32}$$

We construct two matrices A and D that satisfy (16)*, and apply the preconditioning procedure:*

$$\tilde{X} = (X-m_1)A, \quad \tilde{Y} = (Y-m_2)(A^T)^{-1}. \tag{33}$$

If the optimal map between $\tilde{\mu}$ and $\tilde{\nu}$ is $\tilde{Y} = T(\tilde{X})$, the optimal map between $X \sim \mu$ and $Y \sim \nu$ is

$$Y = [m_2 + T((X-m_1)A)A^T]. \tag{34}$$

This preconditioning procedure moves the two given measures into new measures with zero mean and the same diagonal covariance matrix. An extra step that one can add to the preconditioning procedure uses the scaling pairs to normalize both measures so that they have total variance one. In the numerical experiments for this article we do not perform this extra step.

The computational cost of this procedure is low. Assuming we work on datasets with n sample points in p-dimensional space, the computational cost is in the order of np^2+p^3. When $p \ll n$, the main cost is just calculating empirical covariance matrices.

One straightforward theoretical application of the procedure is a simple derivation of the optimal map between multivariate normal distributions. If $X \sim N(m_1, \Sigma_1)$ and $Y \sim N(m_2, \Sigma_2)$, the $\tilde{X}$ and $\tilde{Y}$ resulting from the application of the preconditioning procedure have the same distribution $N(0, D)$. Since the optimal coupling between identical measures is the identity map, the optimal map between $N(m_1, \Sigma_1)$ and $N(m_2, \Sigma_2)$ is

$$Y = m_2 + (X-m_1)AA^T = m_2 + (X-m_1)\Sigma_1^{-1/2}(\Sigma_1^{1/2}\Sigma_2\Sigma_1^{1/2})^{1/2}\Sigma_1^{-1/2}, \tag{35}$$

a result that agrees with the one found in [12] through different means.

This procedure also gives an alternative proof to the following lower bound introduced in [7].

THEOREM 5.2. *Suppose (X,Y) is the optimal coupling between μ and ν. Let $m_1 = \mathbb{E}X$ and $m_2 = \mathbb{E}Y$, and let $\Sigma_{1,2}$ be their respective covariance matrices. Denoting the nuclear norm of a matrix M by $\|M\|_*$, we have the following lower bound for the total transportation cost:*

$$\mathbb{E}\|X-Y\|^2 \geq \|m_1 - m_2\|^2 + \|\Sigma_1\|_* + \|\Sigma_2\|_* - 2\|\Sigma_1^{1/2}\Sigma_2^{1/2}\|_*. \tag{36}$$

Proof. This bound follows directly from the estimation in the proof of Property 3.11. Since

$$\|\Sigma_1\|_* = \mathrm{tr}(\Sigma_1), \quad \|\Sigma_2\|_* = \mathrm{tr}(\Sigma_2), \quad \|\Sigma_1^{1/2}\Sigma_2^{1/2}\|_* = \sum_{i=1}^{N} d_i,$$

applying the *optimal affine pair* to general random variables X and Y, we have

$$\mathbb{E}\|X-Y\|^2 = \|m_1 - m_2\|^2 + \|\Sigma_1\|_* + \|\Sigma_2\|_* - 2\|\Sigma_1^{1/2}\Sigma_2^{1/2}\|_* + \mathbb{E}\|\tilde{X} - \tilde{Y}\|^2.$$

Since clearly $\mathbb{E}\|\tilde{X} - \tilde{Y}\|^2$ is nonnegative, we derive the lower bound (36) along with the condition for the bound to be sharp. ∎

A more general application of this procedure is to precondition measures and datasets before applying numerical optimal transport algorithms.

In practice, instead of continuous probability measures in closed form, one often has only sample points drawn from otherwise unknown distributions. Applying the procedure of this paper to precondition a problem posed in terms of samples is straightforward, since the preconditioning maps act on the random variables, and hence on the sample points. The only difference is that instead of the true mean values and covariance matrices, one uses estimates, such as their empirical counterparts, to define the preconditioning maps.

A natural question is how applying the preconditioning procedure improves the overall algorithm performance.

First, as discussed above, the preconditioning procedure itself is very fast—as it involves only the estimation of means and covariance matrices from the data. This cost is typically negligible in comparison to the cost of any algorithm for numerical optimal transport. With this low computational expense, the procedure can accurately explain and remove a significant portion of the transportation cost. As we will see in the shape transform and color transfer examples below, this portion is substantial in many applications.

The other benefit of the procedure is that it will potentially improve the performance of the numerical optimal transport algorithm. This will be true if the underlying algorithm finds optimal transport problems with closer measures easier to solve. This is because we proved in Property 3.11 that the preconditioned measures are closer to each other than the original ones.

As the preconditioned measures $\tilde{\mu}, \tilde{\nu}$ are closer to each other, the map between them is closer to the identity. Since this map relates to the dual variable ϕ, the dual problem also simplifies. For the standard square-distance cost, the dual problem adopts the form

$$\min_{\phi,\psi} \int \phi(x)\tilde{\mu}(x)dx + \int \psi(y)\tilde{\nu}(y)dy, \quad \phi(x) + \psi(y) \geq x \cdot y.$$

As a guide, consider first the extreme situation where $\tilde{\mu} = \tilde{\nu}$. Then the problem reduces to

$$\min_{\phi,\psi} \int [\phi(x) + \psi(x)]\, \tilde{\mu}(x)dx, \quad \phi(x) + \psi(y) \geq x \cdot y,$$

with solution $\phi(x) = \psi(x) = \frac{1}{2}\|x\|^2$, corresponding to the identity map $T(x) = \nabla\phi(x) = x$. Generally, preconditioning does not make $\tilde{\mu}$ and $\tilde{\nu}$ identical, yet it gives them the same mean $\bar{x}$ and diagonal covariance matrix D. Under these conditions, Property 3.11 shows that among all affine maps, the identity brings the two distributions closest, a result that, translated into the dual variables, states that the optimal quadratic ϕ and ψ are again $\phi(x) = \psi(x) = \frac{1}{2}\|x\|^2$. Hence one can write

$$\phi(x) = \frac{\|x\|^2}{2} + \tilde{\phi}(x), \quad \psi(y) = \frac{\|y\|^2}{2} + \tilde{\psi}(y),$$

which reduces the problem to

$$\min_{\tilde{\phi},\tilde{\psi}} \int \tilde{\phi}(x)\tilde{\mu}(x)dx + \int \tilde{\psi}(y)\tilde{\nu}(y)dy, \quad \tilde{\phi}(x) + \tilde{\psi}(y) \geq -\|x - y\|^2,$$

for which even the simplest choice $\tilde{\phi}(x) = \tilde{\psi}(y) = 0$, which satisfies the constraints trivially, yields the best possible linear map. Thus algorithms based on the dual formulation can take this as a starting point for further improvement of the objective function.

6. Numerical experiments. Our first example concerns optimal transport problems between two-dimensional normal distributions. Consider μ and ν defined by

$$\mu = N\left([1,1], \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}\right), \quad \nu = N\left([-1,0], \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}\right). \tag{37}$$

We generate $N = 200$ data points $\{x_i\}_{i=1}^{200}$ and $\{y_i\}_{i=1}^{200}$ from each distribution. The distributions and sample sets are shown in Figures 1(a) and 1(b).

We then perform the preconditioning procedure on both the distributions and the sample sets. Notice that the two versions should give slightly different results, because in the sample-based version empirical statistics are used instead of the true ones. The results are shown in Figures 1(c) and 1(d). The preconditioning procedure for continuous measures by definition makes $\tilde{\mu} = \tilde{\nu}$. On the other hand, the two preconditioned sample sets are consistent with the preconditioned measures.

In the second example, we test the preconditioning procedure on more complicated distributions. We define both μ and ν to be Gaussian mixtures:

$$\begin{aligned} \mu &= \tfrac{1}{2}N\left([2,-1], \begin{pmatrix} 1/4 & 0 \\ 0 & 1/4 \end{pmatrix}\right) + \tfrac{1}{2}N\left([2,-3], \begin{pmatrix} 1/2 & 1/4 \\ 1/4 & 1/4 \end{pmatrix}\right), \\ \nu &= \tfrac{2}{3}N\left([2,1], \begin{pmatrix} 3 & -1 \\ -1 & 2 \end{pmatrix}\right) + \tfrac{1}{3}N\left([-2,1], \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}\right). \end{aligned} \tag{38}$$

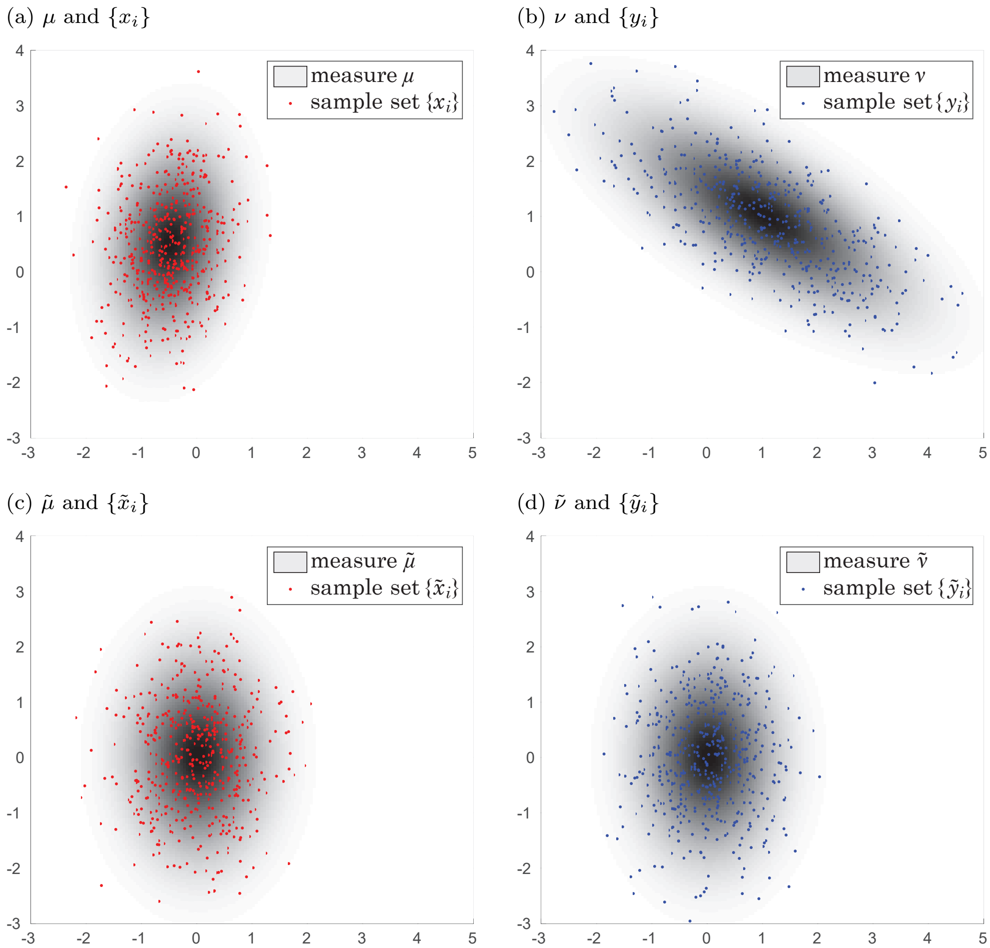

Fig. 1. *Preconditioning on the two Gaussian distributions μ and ν defined in* (37)*. Sample sets $\{x_i\}$ and $\{y_i\}$ are sampled from μ and ν, respectively, each with sample size* 200*. In* (c) and (d)*, the preconditioned measures $\tilde{\mu}$ and $\tilde{\nu}$ are derived from μ and ν by the preconditioning procedure. $\{\tilde{x}_i\}$ and $\{\tilde{y}_i\}$ are transferred from the original sample sets with maps defined by their empirical mean values and covariance matrices.*

In Figures 2(c) and 2(d) the preconditioned datasets have the same diagonal covariance matrix and are closer to each other than in the original datasets. As in the first example, the preconditioned sample sets are consistent with the corresponding preconditioned measures. This shows numerically that the preconditioning procedure on sample sets is consistent with the procedure on continuous measures.

In the third example, we apply the preconditioning procedure along with the sample-based numerical optimal transport algorithm introduced in [14], which takes sample sets as input and compares and transfers them through feature functions. This iterative algorithm approaches the optimal map by gradually approximating the McCann interpolant [16] and updating the local transfer maps. We apply the preconditioning procedure and give the preconditioned sample sets to the algorithm. Then we take the optimal map from the algorithm's output and transform it to solve the original problem. The preconditioning procedure is crucial on two grounds: not

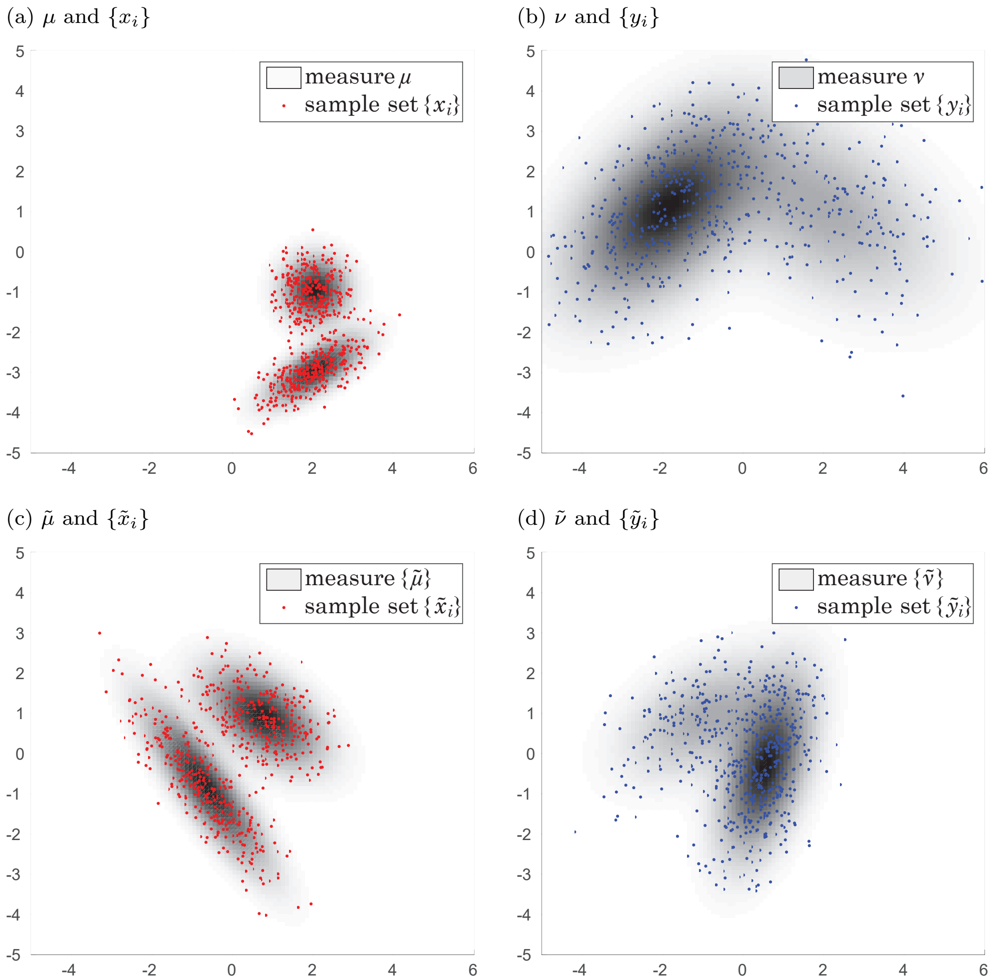

FIG. 2. *The distributions μ and ν are the Gaussian mixtures defined in* (38). *$\{x_i\}$ and $\{y_i\}$ are derived in the same way as in Figure* 1. *In* (c) and (d), *the two preconditioned sample sets $\{\tilde{x}_i\}$ and $\{\tilde{y}_i\}$ are transferred from the original datasets through maps defined in terms of their empirical mean values and covariance matrices.*

only does the algorithm perform better on the preconditioned sample sets, which are closer to each other than the original ones, but feature selection becomes easier, as the same features describe the two distributions at similar levels of precision.

We choose a two-dimensional shape transform problem to test the algorithm. The problem involves finding the optimal transport between two geometrical objects, which can be described in probabilistic terms by introducing a uniform distribution within the support of each. For demonstration purposes, consider the specific task of transforming an ellipse into a ring (Figure 3(a)), described by

$$\begin{aligned} \Omega_2 &= \{(x,y)\,\big|\,1 \leq 3(x-5)^2 + 2(y+1)^2 - (x-5)(y+1) \leq 9\}, \\ \Omega_1 &= \{(x,y)\,\big|\,(x-1)^2 + 10y^2 \leq 1\}. \end{aligned} \tag{39}$$

Both sample sets are drawn from uniform distributions within each region, with the sample size set to 1000 points per sample set.

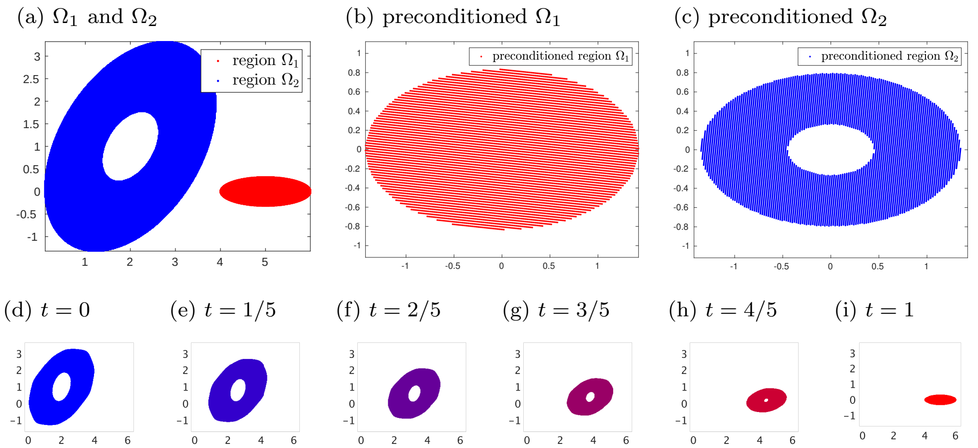

FIG. 3. *Shape transformation problem. The two regions Ω_1 and Ω_2 are shown in* (a)*, and their preconditioned images are shown in* (b) and (c). (d)–(i) *illustrate the McCann interpolation of the optimal map, at times shown in the titles. All computations are carried out on sample sets drawn from the corresponding region. For the plots, we estimate the density function $p(x)$ for each sample set and display the area with $p(x) > \varepsilon$, where ϵ is a small constant. The density functions are estimated by a kernel density estimator with optimal kernel parameters.*

This is a challenging optimal transport problem, since (a) the locations and sizes of the two regions are different; (b) the topological structures of the two regions are different, as one is simply connected and the other is not; (c) both regions have sharp boundaries, which makes the solution singular; and (d) since both shapes are eccentric, the optimal map between them is not essentially one-dimensional as in the transformation between a circle and a circular ring.

The preconditioned regions are shown in Figures 3(b) and 3(c); they share the same mean and diagonal covariance matrix. The two preconditioned regions are much closer to each other, the blue one (b) distinguished by its hole and a slightly smaller radius. Using the sample-based algorithm on the preconditioned sample sets, we find the optimal map T between the two preconditioned regions. Reversing the preconditioning step, the map can then be transformed back to the optimal map between Ω_1 and Ω_2. The map and its McCann interpolation are shown in the second row of Figure 3. Without the preconditioning step, the procedure would have produced much poorer results and at a much higher computational expense.

As a fourth test problem, we apply the preconditioning algorithm to color-transfer problems [21, 27, 10], whose general objective is to recolor an input image so that its colors resemble those of a target image. One can view the set of colors of an image as a distribution and find the optimal map between the input and the target, using the earth mover's distance (EMD) [22] as a quantification of the transfer required.

For the purpose of testing the preconditioning algorithm, we follow the meta-algorithm introduced in [19] without the postprocessing step, since we are concerned here only with the quality of the optimal transport step rather than with the overall quality of the entire algorithm. We first outline the meta-algorithm and refer the reader to [19] for a more detailed description.

Since our focus is on the optimal transport step, we strictly follow all the other steps as described in [19]. For the optimal transport step, we adopt the preconditioning algorithm along with the sample-based algorithm as in the shape transfer problem

Algorithm 1 Color-transfer meta-algorithm.

Sample set formulation: Retrieve the color sample set in the RGB space from each image by gathering the three-dimensional color (RGB) of all the pixels in the image.

Spatio-color clustering: Cluster each color sample set and construct a smaller weighted sample set using a superpixel method [1], which takes into account not only the colors but also the location of the pixels.

Optimal transport: Construct an optimal map $\mathcal{T}(\cdot)$ from the input sample set to the target sample set.

Image synthesis: As the map $\mathcal{T}(\cdot)$ is derived at the superpixel scale, we reconstruct the color map for each pixel from $\mathcal{T}(\cdot)$ using a maximum likelihood estimation [24].

introduced above. The color transfer application fits the optimal transport sample-based setting, since the input consists of sets of data points in the three-dimensional (RGB) space.

We denote the input and target weighted sample sets derived from the space-color clustering step by $\{U_i \in \mathbb{R}^3\}_{i=1}^{N_x}$ with weights $\{\upsilon_i\}_{i=1}^{N_x}$ and $\{V_i \in \mathbb{R}^3\}_{i=1}^{N_y}$ with weights $\{\nu_i\}_{i=1}^{N_y}$, respectively. To demonstrate the advantage of preconditioning, we apply the algorithm in three different ways:

No preconditioning (NP): We apply directly a weighted version of the sample-based algorithm to $\{U_i\}_{i=1}^{N_x}$, $\{V_i\}_{i=1}^{N_y}$ and find the optimal map $\mathcal{T}_1(\cdot)$.

Preconditioning-only (PO): We apply the preconditioning algorithm to the two weighted sample sets, find the preconditioning maps F and G, and directly define as the resulting map their composition $\mathcal{T}_2 = G^{-1} \circ F$.

With preconditioning (P): We first use the preconditioning algorithm to get F and G. Then we apply the sampled-based algorithm to the preconditioned sample sets $\{F(U_i)\}_{i=1}^{N_x}$ and $\{G(V_i)\}_{i=1}^{N_y}$ to find a map $\mathcal{T}'$. Finally, we compute the optimal map consisting of the composition of all three: $\mathcal{T}_3 = G^{-1} \circ \mathcal{T}' \circ F$.

Our objective is to compare P and NP, but we added PO as an additional reference. As proved above, PO will optimally match the mean vectors and covariance matrices of two sample sets, which can be seen as a simple relaxation of the optimal transport problem.

We use two examples from the color-transfer literature. The input and target images are shown in the first two columns of Figure 4. We follow the meta-algorithm, Algorithm 1, with the same parameters used in [19]. Specifically, for the superpixel method we set the number of superpixels to 2000 and the compactness to 2. For the image synthesis step, we set the covariance rescaling parameters to $\sigma_x = \sigma_y = 10$, $\sigma_r = \sigma_g = \sigma_b = 1$.

We can see that the NP and P algorithms do produce different results, as a significant share of the differences between the input and target images is explained and removed by the preconditioning step in the P algorithm. In both examples, while the transformed images from both algorithms are reasonable, the images from the NP algorithm do have some artifacts, while the images from the P algorithm are smoother and contain more details.

Another observation is that the PO algorithm also produces quite reasonable results, despite the fact that it only matches means and covariances. The results from the PO and P algorithms have substantial similarities, which indicates that a large

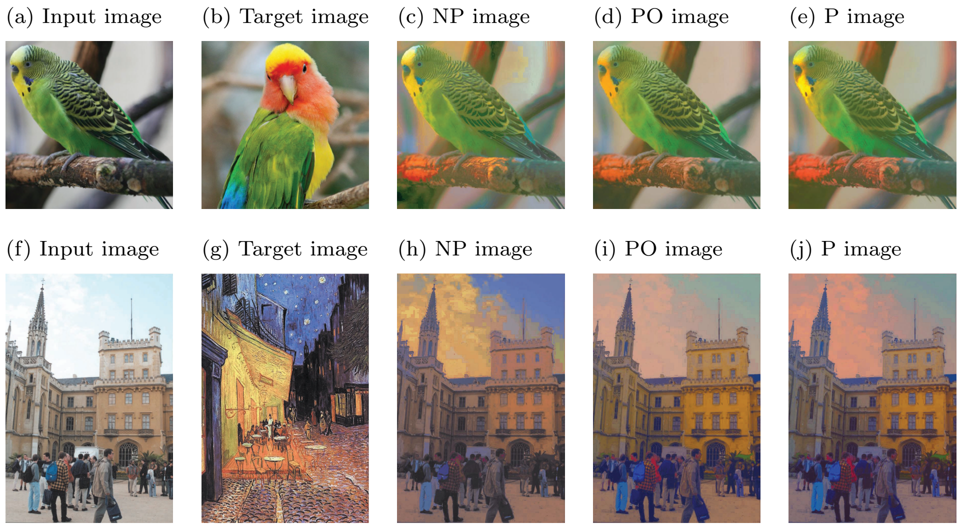

FIG. 4. *Two color-transfer problems. The first column (subfigures* (a) *and* (f)*) contains the input images; the second (*(b) *and* (g)*) contains the target images; the third (*(c) *and* (h)*) displays the images resulting from the* NP *algorithm; the fourth (*(d) *and* (i)*) displays the images resulting from the* PO *algorithm; and the last column (*(e) *and* (j)*) displays the results from the complete algorithm,* P.

portion of the differences (EMD) can be removed by matching means and covariances alone.

7. Conclusions and future work. This paper describes a family of *affine map pairs* that preserves the optimality of transport solutions, and finds an optimal one among them that minimizes the remaining transportation cost. The procedure extends from the square-distance cost to more general cost functions induced by an inner product. Based on these map pairs, we propose a preconditioning procedure which maps input measures or datasets to preconditioned ones while preserving the optimality of the solutions.

The procedure is efficient and easy to implement, and it can significantly reduce the difficulty of the problem in many scenarios. Using this procedure one can directly solve the optimal transport problem between multivariate normal distributions. We tested the procedure both as a stand-alone method and along with a sample-based optimal transport algorithm. The procedure in all cases successfully preconditioned the input measures and datasets, making them more regular and closer to their counterparts.

As future work, one natural extension is to consider nonlinear *admissible map pairs*, which can potentially further reduce the total transportation cost and directly solve a wider class of optimal transport problems. If the family of *admissible map pairs* is rich enough, one can potentially construct a practical optimal transport algorithm from these map pairs alone.

Another possible extension is to the barycenter problem [2]:

$$\min_{\pi_k \in \Pi(\mu_k, \nu), \nu} \sum_{k=1}^{K} w_k \int c(x, y) d\pi_k(x, y), \tag{40}$$

where $\mu_1, \mu_2, \ldots, \mu_K$ are K different measures with positive weights $w_1, w_2, \ldots, w_K$. Instead of the two measures of the regular optimal transport problem, we would like to map K measures simultaneously while preserving the optimality of the solution. The simplest of such maps is the set of translations that give all measures the same mean.

REFERENCES

[1] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, *SLIC superpixels compared to state-of-the-art superpixel methods*, IEEE Trans. Pattern Anal. Mach. Intell., 34 (2012), pp. 2274–2282, https://doi.org/10.1109/TPAMI.2012.120.

[2] M. Agueh and G. Carlier, *Barycenters in the Wasserstein space*, SIAM J. Math. Anal., 43 (2011), pp. 904–924, https://doi.org/10.1137/100805741.

[3] J.-D. Benamou and Y. Brenier, *A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem*, Numer. Math., 84 (2000), pp. 375–393, https://doi.org/10.1007/s002110050002.

[4] Y. Brenier, *Polar factorization and monotone rearrangement of vector-valued functions*, Comm. Pure Appl. Math., 44 (1991), pp. 375–417, https://doi.org/10.1002/cpa.3160440402.

[5] G. Buttazzo, A. Pratelli, and E. Stepanov, *Optimal pricing policies for public transportation networks*, SIAM J. Optim., 16 (2006), pp. 826–853, https://doi.org/10.1137/040619831.

[6] W. Chen and E. G. Tabak, *An adaptive linear programming methodology for data driven optimal transport*, Numer. Math., submitted.

[7] J. A. Cuesta-Albertos, C. Matrán-Bea, and A. Tuero-Diaz, *On lower bounds for the L^2-Wasserstein metric in a Hilbert space*, J. Theoret. Probab., 9 (1996), pp. 263–283, https://doi.org/10.1007/BF02214649.

[8] M. J. P. Cullen and R. J. Purser, *Properties of the Lagrangian semigeostrophic equations*, J. Atmospheric Sci., 46 (1989), pp. 2684–2697, https://doi.org/10.1175/1520-0469(1989)046<2684:POTLSE>2.0.CO;2.

[9] L. C. Evans, *Partial differential equations and Monge-Kantorovich mass transfer*, in Current Developments in Mathematics, 1997 (Cambridge, MA), International Press, Boston, 1999, pp. 65–126, https://doi.org/10.4310/CDM.1997.v1997.n1.a2.

[10] H. S. Faridul, T. Pouli, C. Chamaret, J. Stauder, E. Reinhard, D. Kuzovkin, and A. Trémeau, *Colour mapping: A review of recent methods, extensions and applications*, Comput. Graphics Forum, 35 (2016), pp. 59–88, https://doi.org/10.1111/cgf.12671.

[11] W. Gangbo and R. J. McCann, *Shape recognition via Wasserstein distance*, Quart. Appl. Math., 58 (2000), pp. 705–737.

[12] C. R. Givens and R. M. Shortt, *A class of Wasserstein metrics for probability distributions*, Michigan Math. J., 31 (1984), pp. 231–240, https://doi.org/10.1307/mmj/1029003026.

[13] L. V. Kantorovich, *On a problem of Monge*, J. Math. Sci. (N.Y.), 133 (2006), p. 1383, https://doi.org/10.1007/s10958-006-0050-9.

[14] M. Kuang and E. G. Tabak, *Sample-based optimal transport and barycenter problems*, Comm. Pure Appl. Math., submitted.

[15] J. N. Mather, *Action minimizing invariant measures for positive definite Lagrangian systems*, Math. Z., 207 (1991), pp. 169–207.

[16] R. J. McCann, *A convexity principle for interacting gases*, Adv. Math., 128 (1997), pp. 153–179, https://doi.org/10.1006/aima.1997.1634.

[17] G. Monge, *Mémoire sur la théorie des déblais et des remblais*, De l'Imprimerie Royale, Paris, 1781.

[18] A. M. Oberman and Y. Ruan, *An Efficient Linear Programming Method for Optimal Transportation*, preprint, https://arxiv.org/abs/1509.03668, 2015.

[19] J. Rabin, S. Ferradans, and N. Papadakis, *Adaptive color transfer with relaxed optimal transport*, in Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP), 2014, pp. 4852–4856.

[20] S. T. Rachev and L. Rüschendorf, *Mass Transportation Problems. Volume* I. *Theory*, Springer-Verlag, New York, 1998.

[21] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley, *Color transfer between images*, IEEE Comput. Graph. Appl., 21 (2001), pp. 34–41, https://doi.org/10.1109/38.946629.

[22] Y. Rubner, C. Tomasi, and L. J. Guibas, *A metric for distributions with applications to image databases*, in Proceedings of the Sixth International Conference on Computer Vision, IEEE Computer Society, Washington, DC, 1998, pp. 59–66, https://doi.org/10.1109/ICCV.1998.710701.

[23] J. Solomon, F. de Goes, G. Peyré, M. Cuturi, A. Butscher, A. Nguyen, T. Du, and L. Guibas, *Convolutional Wasserstein distances: Efficient optimal transportation on geometric domains*, ACM Trans. Graph., 34 (2015), 66, https://doi.org/10.1145/2766963.

[24] Y.-W. Tai, J. Jia, and C.-K. Tang, *Local color transfer via probabilistic segmentation by expectation-maximization*, in Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), Vol. 1, IEEE Computer Society, Washington, DC, 2005, pp. 747–754, https://doi.org/10.1109/CVPR.2005.215.

[25] G. Trigila and E. G. Tabak, *Data-driven optimal transport*, Comm. Pure. Appl. Math., 69 (2016), pp. 613–648, https://doi.org/10.1002/cpa.21588.

[26] C. Villani, *Optimal Transport: Old and New*, Grundlehren Math. Wiss. 338, Springer-Verlag, Berlin, 2009, https://doi.org/10.1007/978-3-540-71050-9.

[27] T. Welsh, M. Ashikhmin, and K. Mueller, *Transferring color to greyscale images*, ACM Trans. Graph., 21 (2002), pp. 277–280, https://doi.org/10.1145/566570.566576.