State tracking of linear ensembles via optimal mass transport

Yongxin Chen and Johan Karlsson

Abstract—We consider the problems of tracking an ensemble of indistinguishable agents with linear dynamics based only on output measurements. In this setting, the dynamics of the agents can be modeled by distribution flows in the state space and the measurements correspond to distributions in the output space. In this paper we formulate the corresponding state estimation problem using optimal mass transport theory with prior linear dynamics, and the optimal solution gives an estimate of the state trajectories of the ensemble. For general distributions of systems this can be formulated as a convex optimization problem which is computationally feasible with when the number of state dimensions is low. In the case where the marginal distributions are Gaussian, the problem is reformulated as a semidefinite programming and can be efficiently solved for tracking systems with a large number of states.

I. INTRODUCTION

The optimal mass transport theory provides a geometric framework for mapping a distribution to another one in a way that minimizes the total transport cost [20]. This has been used in many contexts, traditionally for application in economics and logistics [12], and more recently in imaging and machine learning [2], [11], [14], [16] as well as systems and controls [6], [7], [10]. In case when the transport cost is quadratic, the problem may be formulated as a fluid dynamics problem [1]. It can also be viewed as an optimal control problem of the density of the particles that obey dynamics problem [1]. It can also be viewed as an optimal control problem of the density of the particles that obey the dynamics \( \dot{x}(t) = u(t) \) [5]. In the subsequent paper [8] a natural generalization of this problem is introduced where the underlying linear dynamics of the particles become \( \dot{x}(t) = Ax(t) + Bu(t) \).

In this work we consider the extension of this framework to the case where the full state information is not available. In particular we consider the tracking problems where we seek to estimate the states of several identical and indistinguishable systems from only their joint outputs. This is also known as state estimation of ensembles, see [22], [21]. One of the main obstacles is that it is not known which output that is generated by a certain subsystem, hence an association problem has to be solved. A brute force approach to this would result in a combinatorial problem. However, by formulating this as an optimal transport problem which is convex and where the number of variables only grow linearly in the number of stages. Our formulation not only allows for tracking a finite number of particles, but also applies to the more general problems of tracking distributions. We also consider the case where the underlying distributions are Gaussian. In this case the number of variables can be reduced significantly and the problem can be solved efficiently with large number of state dimensions.

At the very high level, we are developing a framework to smoothly interpolate a sequence of probability densities, in a way akin to smoothly interpolate several points in the Euclidean space. Indeed, the cubic spline interpolation of several points has a variational formulation in the flavor of optimal control [15]. The cubic spline counterpart in the space of distributions has been recently studied in [4], [3]. Our work can be viewed as a generalization of these where more general underlying linear dynamics, instead of simple second order integrator, are considered.

The rest of the paper is structured as follows. Section II is a brief introduction to the optimal mass transport theory. In Section III we formulate the tracking problems using optimal mass transport theory. The case where the densities are assumed Gaussian is discussed in Section IV. Finally, we present several numerical examples in Section V to illustrate our framework.

II. BACKGROUND ON OPTIMAL MASS TRANSPORT

Monge’s original formulation of optimal mass transport is as follows (see, e.g., [20]). Consider two nonnegative distributions, \( \rho_0 \) and \( \rho_1 \), of the same mass, defined on a compact set \( \mathcal{X} \subset \mathbb{R}^n \). The optimal mass transport problem seeks a transport function \( f : \mathcal{X} \rightarrow \mathcal{X} \) that minimizes the transportation cost

\[
\int_{\mathcal{X}} c(x, f(x))\rho_0(x)dx
\]

over all the mass preserving maps from \( \rho_0 \) to \( \rho_1 \), namely,

\[
\int_{x \in \mathcal{U}} \rho_1(x)dx = \int_{f(x) \in \mathcal{U}} \rho_0(x)dx \quad \text{for all } \mathcal{U} \subset \mathcal{X}, \tag{1}
\]

which is often denoted \( f\#\rho_0 = \rho_1 \). The function \( c(x_0, x_1) : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}_+ \) is a cost function that describes the cost for transporting a unit mass from \( x_0 \) to \( x_1 \). The Monge’s problem is usually difficult to solve due to the nontrivial constraint \( f\#\rho_0 = \rho_1 \) and the minimum may not exist. To overcome these difficulties, Kantorovich proposed a linear programming relaxation

\[
\min_{\pi \in \Pi(\rho_0, \rho_1)} \int_{\mathcal{X} \times \mathcal{X}} c(x_0, x_1)d\pi(x_0, x_1) \tag{2}
\]
where $\Pi(\rho_0, \rho_1)$ denotes the set of all joint distributions between $\rho_0$ and $\rho_1$. In fact, when $\rho_0$ and $\rho_1$ are absolutely continuous, these two formulations are equivalent.

When the cost function is quadratic, i.e., $c(x_0, x_1) = \|x_0 - x_1\|^2$, the optimal mass transport problem can be set up as an optimal control problems in fluid dynamics [1]

$$
\begin{align*}
&\min_{u, \rho} \quad \int_0^1 \int_{x \in X} \|u(t, x)\|^2 \hat{\rho}(t, x) dx dt \\
&\text{subject to} \quad \frac{\partial \hat{\rho}}{\partial t} + \nabla \cdot (u \hat{\rho}) = 0 \\
&\quad \rho_k = \hat{\rho}(k, \cdot), \quad \text{for } k = 0, 1.
\end{align*}
$$

This can be interpreted as an optimization problem where the mass distributions represented by infinite-decimal particles, each carrying a cost corresponding to the optimal control problem

$$
\begin{align*}
&\min_u \quad \int_0^1 \|u(t)\|^2 dt \\
&\text{subject to} \quad \dot{x}(t) = u(t), \\
&\quad x(0) = x_0 \text{ and } x(1) = x_1.
\end{align*}
$$

where $x_0, x_1$ are the initial and final position of the particle, respectively. Hence, choosing the quadratic cost in the optimal mass transport problem can be seen as assuming the underlying dynamic being $\dot{x}(t) = u(t)$.

In [8] this was generalized through replacing the cost function that reflects deviation from the trajectory of the underlying system dynamics. It is associated with the linear dynamic

$$
\dot{x}(t) = Ax(t) + Bu(t)
$$

and optimal control problem

$$
\begin{align*}
&\min_u \quad \int_0^1 \|u(t)\|^2 dt \\
&\text{subject to} \quad \dot{x}(t) = Ax(t) + Bu(t), \\
&\quad x(0) = x_0 \text{ and } x(1) = x_1.
\end{align*}
$$

The cost is then given by

$$
c(x_0, x_1) = (x_1 - \Phi x_0)^T Q (x_1 - \Phi x_0).
$$

where $\Phi = e^{A}, Q = M_0^{-1}$, and

$$
M_0 = \int_0^1 e^{A(1-\tau)} BB^T e^{A^T(1-\tau)} d\tau
$$

is the controllability Grammian. Apparently, it reduces to the standard cost $c(x_0, x_1) = \|x_0 - x_1\|^2$ when $A = 0, B = 1$.

III. TRACKING WITH OPTIMAL MASS TRANSPORT FROM OUTPUT MEASUREMENTS

We next extend this connection between optimal transport and optimal control to systems with output measurements. To this end, assume that the underlying dynamic and output measurements corresponds to the linear system

$$
\begin{align*}
\dot{x}(t) &= Ax(t) + Bu(t) \\
y(t) &= Cx(t)
\end{align*}
$$

where $A \in \mathbb{R}^{n \times n}, B \in \mathbb{R}^{n \times p}, C \in \mathbb{R}^{m \times n}$ and $(A, B)$ is a controllable pair. We seek to track the time varying distribution $\hat{\rho}_i$, where each particle abides by (5a), based on the output distributions $\rho_k = C\hat{\rho}_k$ at the times $k = 0, 1, \ldots, T$. Note that only the distribution of the output is available. We don’t have access to the information about each particle, namely, the particles are indistinguishable. For example, with a finite ensemble $(x_1(t), \ldots, x_N(t))$ then $\hat{\rho}_i(t) = \sum_{i=1}^N \delta(x - x_i(t)) \in \mathcal{M}_+(\mathbb{R}^n)$ for $t \in [0, T]$ and the outputs are $\rho_k = \sum_{i=1}^N \delta(y - Cx_i(k)) \in \mathcal{M}_+(\mathbb{R}^m)$ for $k = 0, 1, \ldots, T$, where $\delta$ is the Dirac delta function and $\mathcal{M}_+(\mathbb{R}^n)$ is the set of measures on $\mathbb{R}^n$. For determining identifiability of this problem, see [22].

We propose to model this tracking problem as the following optimal mass transport problem. We seek a flow of nonnegative measures $\hat{\rho} : t \mapsto \mathcal{M}_+(X)$ that minimize

$$
\begin{align*}
&\min_{u, \hat{\rho}} \quad \int_{t=0}^T \int_{x \in X} \|u(t, x)\|^2 \hat{\rho}(t, x) dx dt \\
&\text{subject to} \quad \frac{\partial \hat{\rho}}{\partial t} + \nabla \cdot ((Ax + Bu)\hat{\rho}) = 0 \\
&\quad \rho_k = C\hat{\rho}(k, \cdot), \quad \text{for } k = 0, 1, \ldots, T.
\end{align*}
$$

Reformulating this using the Kantorovich formulation of the optimal transport problems we arrive at the linear programming problem

$$
\begin{align*}
&\min_{\hat{\rho} \in \mathcal{M}_+(X^2)} \quad \sum_{k=0}^{T-1} \int_{(x_k, x_{k+1}) \in X^2} c(x_k, x_{k+1}) d\pi_k(x_k, x_{k+1}) \\
&\text{subject to} \quad \int_{x_{k+1} \in X} d\pi_k(x_k, x_{k+1}) = d\hat{\rho}_k(x_k) \\
&\quad \int_{x_k \in X} d\pi_k(x_k, x_{k+1}) = d\hat{\rho}_{k+1}(x_{k+1}) \\
&\quad \rho_k = C\hat{\rho}_k, \quad \text{for } k = 0, 1, \ldots, T
\end{align*}
$$

where the cost function $c$ is given by (4). This optimization problem has a dual formulation in terms of continuous test functions on $C(X)$, denoted $\text{Cont}(C(X))$.

**Theorem 1:** The optimization problem (7) is the dual of

$$
\begin{align*}
&\max_{\phi_k \in \text{Cont}(C(X))} \quad \sum_{k=0}^T \int_{C(X)} \phi_k(y_k) d\rho_k(y_k) \\
&\text{subject to} \quad \sum_{k=0}^T \phi_k(Cx_k) \leq \sum_{k=1}^T c(x_{k-1}, x_k) \\
&\quad \text{for all } x_k \in \mathbb{R}^n, \ k = 0, \ldots, T,
\end{align*}
$$

and the duality gap is zero.

**Proof:** See the appendix for a proof.
When the number of states is large we might therefore want of K-means clustering, as proposed in [21]. An alternative approach would be testing all associations between measurements and particles, but this would require a combinatorial search which grows exponentially in $T$. Yet another possible approach for addressing this problem is to use heuristics along the lines of K-means clustering, as proposed in [21].

When the number of states is large we might therefore want to restrict the distributions to certain classes. One such class of particular interest is the Gaussian distributions.

IV. GAUSSIAN CASES

In this section, we zoom in to the case when all the marginal distributions are Gaussian. We assume that, for all $k = 0,1,\ldots,T$, the $k$-th marginal $\rho_k$ of the measurement is a Gaussian distribution with mean $\mu_k$ and covariance $\Sigma_k$. By linearity, the output density tracking problem can be divided into two parts: interpolating the means $\{\mu_k\}$ and interpolating the covariances $\{\Sigma_k\}$.

Interpolating the means is equivalent to solving (6) for a single particle. It reduces to the optimal control problem

\begin{equation}
\min_u \int_0^T \|u(t)\|^2 dt \quad (9a)
\end{equation}

subject to
\begin{equation}
\dot{x}(t) = Ax(t) + Bu(t), \quad 0 \leq t \leq T \quad (9b)
\end{equation}

\begin{equation}
Cx(k) = \mu_k, \quad k = 0,1,\ldots,T. \quad (9c)
\end{equation}

By introducing a Lagrangian multiplier $\lambda(\cdot)$, it is easy to see the optimal control is of the form $u(t) = B^T \lambda(t)$ with $\lambda$ satisfying the dual dynamics $\dot{\lambda}(t) = -A^T \lambda(t)$ for each interval $t \in (k,k+1)$. For each interval, if we fix $x(k), x(k+1)$, then we can obtain a closed-form expression for the optimal cost, which is

$$(x(k+1) - \Phi x(k))^T Q (x(k+1) - \Phi x(k)).$$

Therefore, a strategy to solve (9) is first minimizing $u$ over fixed $x(0),\ldots,x(T)$ and then minimizing the result over $x(0),\ldots,x(T)$.

The covariances part is solved using semidefinite programming (SDP). We first minimize the cost with fixed state $x(0),x(1),\ldots,x(T)$ and then minimize the resulting cost over $x(k), k = 0,1,\ldots,T$ subject to the constraint that $Cx(k)$ is zero mean Gaussian distribution with covariance $\Sigma_k$. For fixed $x(k), k = 0,1,\ldots,T$, the minimum of the cost is given in the quadratic form

$$
\sum_{k=0}^{T-1} c(x(k), x(k+1)) = \sum_{k=0}^{T-1} (x(k+1) - \Phi x(k))^T Q (x(k+1) - \Phi x(k)).
$$

We then minimize this cost subject to the distribution constraint of the output, which reads as

$$
\min \mathbb{E}\{\sum_{k=0}^{T-1} c(x(k), x(k+1))\} \quad (10a)
$$

$$
y(k) = Cx(k) \sim \Sigma_k, \quad k = 0,1,\ldots,T. \quad (10b)
$$

We notice that in the state space, the problem can be viewed as $T$ separate optimal transport problems. However, these problems are coupled through the constraints on the output. Since the cost function is quadratic, it is not difficult to show that the solution remains Gaussian. Thus, the cost becomes

$$
\sum_{k=0}^{T-1} \text{Tr}(Q \hat{\Sigma}_{k+1} + \Phi^T \Phi \hat{\Sigma}_k - 2Q \Phi S_{k,k+1}),
$$

where $S_{k,k+1} = \mathbb{E}\{x(k)x(k+1)^T\}$.

Theorem 2: The density tracking Problem (6) for Gaussian marginals with covariances $\{\Sigma_0, \Sigma_1, \ldots, \Sigma_T\}$ has the SDP formulation

\begin{equation}
\min_{\Sigma,S} T-1 \sum_{k=0}^{T-1} \text{Tr}(Q \hat{\Sigma}_{k+1} + \Phi^T \Phi \hat{\Sigma}_k - 2Q \Phi S_{k,k+1}) \quad (11a)
\end{equation}

subject to
\begin{equation}
\left[
\begin{array}{cc}
\hat{\Sigma}_k & S_{k,k+1} \\
S_{k,k+1}^T & \Sigma_{k+1}
\end{array}
\right] \succeq 0, \quad k = 0,1,\ldots,T-1 \quad (11b)
\end{equation}

\begin{equation}
C \hat{\Sigma}_k C^T = \Sigma_k, \quad k = 0,1,\ldots,T. \quad (11c)
\end{equation}

We remark that it suffice to have constraint (11b) to guarantee a well-defined covariance matrix for the random vector $(x_0,x_1,\ldots,x_T)^T$. This can be proven constructively using graphical models, see [4].

Alternatively, to solve (10), one can choose to minimize the cost over fixed $y(k), k = 0,1,\ldots,T$ first and then minimize result over the distribution of $y(k)$. Let $c_y(y(0),y(1),\ldots,y(T))$ be the minimum of the cost for fixed outputs. More precisely,

$$
c_y(y(0),\ldots,y(T)) = \min_x \sum_{k=0}^{T-1} c(x(k), x(k+1))$$

where $y(k) = Cx(k), \quad k = 0,1,\ldots,T$.

Since $c(\cdot, \cdot)$ is quadratic, $c_y$ is a quadratic function of $y = [y(0), y(1), \ldots, y(T)]^T$. Furthermore, it has the form

$$
c_y(y) = y^T R y
$$

for some positive semidefinite matrix $R$. Thus, problem (10) becomes

\begin{equation}
\min_y \mathbb{E}\{\text{tr}(R y y^T)\} \quad (12a)
\end{equation}

\begin{equation}
y(k) \sim \Sigma_k, \quad k = 0,1,\ldots,T, \quad (12b)
\end{equation}
Fig. 1: Ensemble outputs in example with \( N = 7 \) systems. System dynamic given by (14) and \( \sigma = 0.1 \). Blue trajectory: example of a system trajectory with \( \sigma = 0 \).

or equivalently,

\[
\begin{align*}
\min_{\Sigma_y \geq 0} & \quad \text{tr}(R \Sigma_y) \\
\Sigma_y(k,k) &= \Sigma_k, \quad k = 0, \ldots, T.
\end{align*}
\] (13a)

(13b)

Remark 3: The two SDP formulations (11) and (13) represent two different ways of viewing the density tracking problem. In (11) we lift the marginals distributions to the state space and solve \( T \) separate optimal transport problems. In contrast, (13) approaches the problem directly by considering the joint distributions of \( y(0), y(1), \ldots, y(T) \). In terms of complexity, (11) grows linearly as the number of time points \( T \), while (13) grows as \( T^6 \) in the worse case. Therefore, (11) is better for computational purpose.

After obtaining the marginal covariances \( \{ \hat{\Sigma}_k \} \) for the state variables, we can recover \( \bar{\Sigma}(t), k \leq t \leq k+1 \) from \( \hat{\Sigma}_k, \hat{\Sigma}_{k+1} \) in closed-form for each \( k = 0, \ldots, T - 1 \), using optimal mass transport theory over linear dynamics [8]. It follows that the trajectory of the covariances of the output is given by \( \Sigma(t) = C \Sigma(t) C^T \). Finally, let \( \mu(t) = x(t) \) be the solution to (9), we obtain our optimal density flow being a flow of Gaussian distributions with mean \( \mu(t) \) and covariance \( \Sigma(t) \).

V. EXAMPLES

Two examples are provided to illustrate our framework. In the first example, we explain the tracking of finite number of particles and in the second one, we consider a Gaussian distributions tracking problem. The implementations are made in Matlab using the package CVX, which is a package for specifying and solving convex programs [13], [9].

A. Tracking of particles

We illustrate the tracking of a series of systems with a given system dynamic. Consider the tracking of \( N \) systems with oscillatory dynamics

\[
\begin{align*}
\dot{x}(t) &= Ax(t)dt + \sigma dw(t) \\
y(t) &= Cx(t)
\end{align*}
\]

where the state dynamics is given by

\[
A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 \end{bmatrix}
\]

and where \( dw \) is normalized white Gaussian noise. We seek to recover the full state information of the systems based on only the unordered outputs, observed at the time point \( t = 0, 1, \ldots, 5 \). The problem is solved in the two following examples using the formulation (7), where the state space is discretized. The discretization of the state space at time \( k \) is
given by
\[ \{ (x_1, x_2) | x_1 \in \text{supp}(\rho_k), x_2 \in \text{linspace}([-7, 7], 150) \}, \]
where \text{linspace}([a, b], N) denotes the set of \( N \) uniformly spaced points in the interval \([a, b]\).

For the first example we let the number of particles be \( N = 7 \) and the noise level \( \sigma = 0.1 \). Figure 1 shows the output measurement and an example of a noiseless trajectory corresponding to (14). We seek to group the measurements corresponding to which system they belong. Figure 2 shows the reconstruction based on the optimization problem (7), and both the state estimates correspond well to the true states. Even though there are a few error in the associations, these are between points that are spaced closely together and does not seem to impact the state estimation significantly. Next, we consider an example with \( N = 5 \) particles and noise level \( \sigma = 0.5 \). Figure 3 shows the reconstruction based on the optimization problem (7). Even though the noise level is fairly large, we are able to achieve good reconstructions of the states.

### B. Tracking of Gaussian distributions

We consider a dynamical system consisting of a simple, possibly high dimensional, first order integrator. The dynamics are governed by
\[
A = \begin{bmatrix} 0 & I \\ 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ I \end{bmatrix}, \quad C = \begin{bmatrix} I & 0 \end{bmatrix},
\]
where \( I \) is an identity matrix of proper size \( n \). When the dimension of the output is \( n = 1 \), we randomly generate several covariances and interpolate them using (11). The results are depicted in Figure 4 for \( T = 5, 10 \). It can be observed that the interpolated curves are smooth. Similar results hold for high dimensional setting. Figure 5 depicts several Gaussian distributions with different means and covariances we want to track. The tracking result is shown in Figure 6, which is a natural and smooth interpolation of the observations.

### VI. Conclusions

A framework of tracking the states of indistinguishable particles with linear dynamics using output measures is presented. The measurements are the distributions of the output at several time points. Our framework relies on a recent development of optimal mass transport theory with prior dynamics [8]. In the special case with Gaussian marginals, our problem has a SDP formulation and can be solved efficiently. Developing fast algorithms for the general cases will be a future research topic.

### APPENDIX

#### A. Proof of Theorem 1

The Lagrangian of the problem (8), where the constraint (8b) is relaxed, is given by
\[
L(\{\phi_k\}_{k=0}^T, M) = \sum_{k=0}^T \int_{C(X)} \phi_k(y_k) d\rho_k(y_k)
\]
\[
+ \int_{X^{T+1}} \left( \sum_{k=1}^T c(x_{k-1}, x_k) - \sum_{k=0}^T \phi_k(Cx_k) \right) dM(x)
\]
where $x = (x_0, \ldots, x_T) \in X^{T+1}$ [17, Chapter 8]. Note that since the constraint (8c) is defined by continuous functions, then the dual variable is a nonnegative measure $M \in \mathcal{M}_+(X^{T+1})$. Denote the one dimensional and two-dimensional marginals of $M$ by

$$d\hat{\rho}_k(x_k) = \int_{x_0, \ldots, x_{k-1}, x_{k+1}, \ldots, x_T \in X} dM(x), \quad (15a)$$

$$d\pi_k(x_k, x_{k+1}) = \int_{x_0, \ldots, x_{k-1}, x_{k+2}, \ldots, x_T \in X} dM(x). \quad (15b)$$

For a fixed $M$, the maximum of $L(\{\phi_k\}_{k=0}^T, M)$ with respect to $\phi_k$ is finite only if

$$\int_{C(X)} \phi_k(y_k) d\rho_k(y_k) = \int_X \phi_k(Cx_k) \hat{\rho}_k(x_k)$$

for any continuous functional $\phi_k : C(X) \mapsto \mathbb{R}$. From the Riesz representation theorem (see e.g. [17]) and the definition (11) of mass preserving maps this is equivalent with $C_\# \rho_k = \hat{\rho}_k$. Also, using (15b) the remaining second term of $L(\{\phi_k\}_{k=0}^T, M)$ can be written as

$$\sum_{k=0}^{T-1} \int_{(x_k, x_{k+1}) \in X \times X} c(x_k, x_{k+1}) d\pi_k(x_k, x_{k+1}). \quad (16)$$

Next, note that for any $M \in \mathcal{M}_+(X^{T+1})$ the marginals (15) satisfy the constraints (7b) and (7c) in the optimization problem (7). Conversely we would like to show that for any set of marginals satisfying the constraints (7b)–(7c) for $k = 0, 1, \ldots, T − 1$ there is distribution $M \in \mathcal{M}_+(X^{T+1})$ that satisfies (15). Thus, let the one-dimensional marginals $\hat{\rho}(x_k)$ and two-dimensional marginals $\pi_k(x_k, x_{k+1})$ satisfy (7b) and (7c) for $k = 0, 1, \ldots, T − 1$. From equations (7b)–(7c) it follows that the total mass $\hat{\rho}_k(X) = k$ is the same for all $k = 0, 1, \ldots, T$. Normalizing the total mass to 1, we may consider the Markov chain with transition probabilities from time $k$ to $k + 1$ specified by the joint distribution $\pi(x_k, x_{k+1})/k$. Consequently if we let $M/k$ be the measure corresponding to the joint probabilities of the Markov chain, then $M$ satisfies (15) (see e.g., [19]).

To summarize, we have shown that the dual functional of (8) is given by

$$\max_{\{\phi_k\}_{k=0}^T} L(\{\phi_k\}_{k=0}^T, M) = \begin{cases} 16 & \text{if } d\pi_k \text{ satisfies (7d)} \\ \infty & \text{otherwise} \end{cases}$$

where $d\pi_k$ are the marginals given by (15b). The functions $d\hat{\rho}_k$ and $d\pi_k$ are marginals (15) for some $M \in \mathcal{M}_+(X^{T+1})$ if and only if the marginals satisfies the constraints (7b)–(7c), hence the dual of (8) is given by (7). Furthermore, since (8) has a feasible interior point, the duality gap is zero [17, p. 217].

REFERENCES

[1] J.-D. Benamou and Y. Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. Numerische Mathematik, 84(3):375–393, 2000.

[2] J.-D. Benamou, G. Carlier, M. Cuturi, L. Nenna, and G. Peyré. Iterative Bregman projections for regularized transportation problems. SIAM Journal on Scientific Computing, 37(2):A1111–A1138, 2015.

[3] J.-D. Benamou, T. Gallouët, and F.-X. Vialard. Second order models for optimal transport and cubic splines on the Wasserstein space. arXiv preprint arXiv:1801.04144, 2018.

[4] Y. Chen, G. Conforti, and T.T. Georgiou. Measure-valued spline curves: an optimal transport viewpoint. arXiv preprint arXiv:1801.03186, 2018.

[5] Y. Chen, T.T. Georgiou, and M. Pavon. On the relation between optimal transport and Schrödinger bridges: A stochastic control viewpoint. Journal of Optimization Theory and Applications, 169(2):671–691, 2016.

[6] Y. Chen, T.T. Georgiou, and M. Pavon. Optimal steering of a linear stochastic system to a final probability distribution, Part I. IEEE Trans. on Automatic Control, 61(5):1158–1169, 2016.

[7] Y. Chen, T.T. Georgiou, and M. Pavon. Optimal steering of a linear stochastic system to a final probability distribution, Part III. arXiv:1608.03622, IEEE Trans. on Automatic Control, to appear, 2017.

[8] Y. Chen, T.T. Georgiou, and M. Pavon. Optimal transport over a linear dynamical system. IEEE Transactions on Automatic Control, 62(5):2137–2152, 2017.

[9] Inc. CVX Research. CVX: Matlab software for disciplined convex programming, version 2.0, http://cvxr.com/cvx August 2012.

[10] F. Elvander, A. Jakobsson, and J. Karlsson. Interpolation and extrapolation of Toeplitz matrices via optimal mass transport. arXiv preprint arXiv:1711.03890, 2017.

[11] B. Engquist and B.D. Froese. Application of the Wasserstein metric to seismic signals. Communications in Mathematical Sciences, 12(5), 2014.

[12] A. Galichon. Optimal Transport Methods in Economics. Princeton University Press, 2016.

[13] M.C. Grant and S.P. Boyd. Graph implementations for nonsmooth convex programs. In Recent advances in learning and control, pages 95–110. Springer, 2008.

[14] S. Haker, L. Zhu, A. Tannenbaum, and S. Angenent. Optimal mass transport for registration and warping. International Journal of computer vision, 60(3):225–240, 2004.

[15] J.C. Holladay. A smoothest curve approximation. Mathematical tables and other aids to computation, 11(60):233–243, 1957.

[16] J. Karlsson and A. Ringh. Generalized Sinkhorn iterations for regularizing inverse problems using optimal mass transport. SIAM Journal on Imaging Sciences, 10(4):1935–1962, 2017.

[17] D.G. Luenberger. Optimization by vector space methods. John Wiley & Sons, 1997.

[18] B. Pass. Uniqueness and Monge solutions in the multimarginal optimal transportation problem. SIAM Journal on Mathematical Analysis, 43(6):2758–2775, 2011.

[19] M. Sharpe. General theory of Markov processes, volume 133. Academic press, 1988.

[20] C. Villani. Topics in Optimal Transportation, volume 58. Graduate studies in Mathematics, AMS, 2003.

[21] S. Zeng, H. Ishii, and F. Allgöwer. Sampled observability and state estimation of linear discrete ensembles. IEEE Transactions on Automatic Control, 62(5):2406–2418, 2017.

[22] S. Zeng, S. Waldherr, C. Ebenbauer, and F. Allgöwer. Ensemble observability of linear systems. IEEE Transactions on Automatic Control, 61(6):1452–1465, 2016.