23rd International Symposium on Mathematical Theory of Networks and Systems
Hong Kong University of Science and Technology, Hong Kong, July 16-20, 2018

Tracking distributions of linear dynamical systems: an optimal mass transport approach

Yongxin Chen and Johan Karlsson

Abstract— **We consider the problems of tracking a set of indistinguishable agents with linear dynamics based only on output measurements. The behaviors of the agents may be modeled by distributions. We formulate the problems using optimal mass transport theory with prior linear dynamics. Though our problem has a convex formulation, with general purpose solver, it is only computationally feasible if the state dimension 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 large state dimension.**

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 [16]. 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], [13], [15] 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 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 extended abstract, based on [9], 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 [18], [17]. 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 the number of variables only grow linearly in the number of states. 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 when 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 [14]. 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.

*This work was in part supported by the Swedish Research Council (VR) grant 2014-5870 and the Swedish Foundation of Strategic Research (SSF) grant AM13-0049.

Y. Chen is with the Department of Electrical and Computer Engineering, Iowa State University, email:yongchen@iastate.edu. J. Karlsson is with Department of mathematics, KTH Royal Institute of Technology. email: johan.karlsson@math.kth.se.

II. Background on optimal mass transport

Monge's original formulation of optimal mass transport is as follows (see, e.g., [16]). Consider two nonnegative distributions, ρ_0 and ρ_1, of the same mass, defined on a set $X \subset \mathbb{R}^n$. The optimal mass transport problem seeks a transport function $f : X \to X$ that minimizes the transportation cost

$$\int_X c(x, f(x))\rho_0(x)dx$$

over all the mass preserving maps from ρ_0 to ρ_1, namely,

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

which is often denoted $f_\# \rho_0 = \rho_1$. The function $c(x_0, x_1) : X \times X \to \mathbb{R}_+$ 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 $f_\# \rho_0 = \rho_1$. To overcome this difficulty, Kantorovich proposed a linear programming relaxation

$$\min_{\pi\in\Pi(\rho_0,\rho_1)} \int_{X\times X} c(x, y)d\Pi(x, y) \tag{1}$$

where $\Pi(\rho_0, \rho_1)$ denotes the set of all joint distributions between ρ_0 and ρ_1. In fact, when ρ_0 and ρ_1 are absolutely continuous, these two formulations are equivalent.

When the cost function is quadratic, i.e., $c(x_0, x_1) = \|x_0 - x_1\|_2^2$, the optimal mass transport problem can be set up as an optimal control problems in fluid dynamics [1]

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

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

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

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

$$\dot{x}(t) \quad = \quad Ax(t) + Bu(t) \tag{2}$$

and optimal control problem

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

The cost is then given by

$$c(x_0, x_1) = (x_1 - \Phi x_0)^T Q (x_1 - \Phi x_0). \tag{3}$$

where $\Phi = e^A, Q = M_{10}^{-1}$, and

$$M_{10} = \int_0^1 e^{A(1-\tau)} BB^T e^{A^T(1-\tau)} d\tau$$

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

$$\begin{aligned}\dot{x}(t) &= Ax(t) + Bu(t) && \text{(4a)} \\ y(t) &= Cx(t) && \text{(4b)}\end{aligned}$$

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 $\hat{\rho}(t)$, where each particle abides by (4a), based on the output distributions $\rho_k = C_{\#}\hat{\rho}(t)$ 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 determining identifiability of this problem, see [18].

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

$$\begin{aligned}\min_{u,\hat{\rho}} \quad & \int_{t=0}^T \int_{x\in X} \|u(t,x)\|^2 \hat{\rho}(t,x) dx dt && \text{(5a)} \\ \text{subject to} \quad & \frac{\partial \hat{\rho}}{\partial t} + \nabla \cdot ((Ax + Bu)\hat{\rho}) = 0 && \text{(5b)} \\ & \rho_k = C_{\#}\hat{\rho}(k, \cdot), \quad \text{for } k = 0, 1, \ldots, T. && \text{(5c)}\end{aligned}$$

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

$$\begin{aligned}\min_{\pi_k \in \mathcal{M}_+(X\times X)} \quad & \sum_{t=0}^{T-1} \int_{(x,y)\in X\times X} c(x,y) d\pi_k(x,y) && \text{(6a)} \\ \text{subject to} \quad & \int_{y\in X} d\pi_k(x,y) = d\hat{\rho}_k(x) && \text{(6b)} \\ & \int_{x\in X} d\pi_k(x,y) = d\hat{\rho}_{k+1}(y) && \text{(6c)} \\ & \rho_k = C_{\#}\hat{\rho}_k, \quad \text{for } k = 0, 1, \ldots, T, && \text{(6d)}\end{aligned}$$

where the cost $c(x_0, x_1)$ is given by (3).

This is a linear programming problem that can be solved using standard methods if the number of states are small. However, it suffers from the curse of dimensionality and when the number of states are large we need to restrict the distributions to certain classes. One such class of particular interest is the Gaussian distributions.

IV. Gaussian cases

In this section, we focus on the case when all the marginal distributions are Gaussian. We assume that, for all $k = 0, 1, \ldots, T$, the k-th marginal ρ_k of the measurement is a Gaussian distribution with mean μ_k and covariance Σ_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 (5) for a single particle. It reduces to the optimal control problem

$$\begin{aligned}\min_{u} \quad & \int_0^T \|u(t)\|^2 dt && \text{(7a)} \\ \text{subject to} \quad & \dot{x}(t) = Ax(t) + Bu(t), \quad 0 \le t \le T && \text{(7b)} \\ & Cx(k) = \mu_k, \qquad k = 0, \ldots, T. && \text{(7c)}\end{aligned}$$

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 λ 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 (7) 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 $x(0), x(1), \dots, x(T)$ and then minimize the resulting cost over $x(k), k = 0, 1, \dots, T$ subject to the constraint that $Cx(k)$ is zero mean Gaussian distribution with covariance Σ_k. For fixed $x(k), k = 0, 1, \dots, T$, 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

$$\min \quad \mathbb{E}\{\sum_{k=0}^{T-1} c(x(k), x(k+1))\} \tag{8a}$$
$$y(k) = Cx(k) \sim \Sigma_k, \; k = 0, 1, \dots, T. \tag{8b}$$

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

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

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

Theorem 1: The density tracking Problem (5) for Gaussian marginals with covariances $\{\Sigma_0, \Sigma_1, \dots, \Sigma_T\}$ has the SDP formulation

$$\min_{\hat{\Sigma}, S} \sum_{k=0}^{T-1} \mathrm{Tr}(Q\hat{\Sigma}_{k+1} + \Phi^T Q \Phi \hat{\Sigma}_k - 2Q\Phi S_{k,k+1}) \tag{9a}$$
$$\begin{bmatrix} \hat{\Sigma}_k & S_{k,k+1} \\ S_{k,k+1}^T & \hat{\Sigma}_{k+1} \end{bmatrix} \geq 0, \quad k = 0, \dots, T-1 \tag{9b}$$
$$C\hat{\Sigma}_k C^T = \Sigma_k, \quad k = 0, 1, \dots, T. \tag{9c}$$

We remark that it suffice to have constraint (9b) to guarantee a well-defined covariance matrix for the random vector $(x_0, x_1, \dots, x_T)^T$. This can be proven constructively using graphical models, see [4].

After obtaining the marginal covariances $\{\hat{\Sigma}_k\}$ for the state variables, we can recover $\hat{\Sigma}(t), k \leq t \leq k+1$ from $\hat{\Sigma}_k, \hat{\Sigma}_{k+1}$ in closed-form for each $k = 0, \dots, T-1$, using optimal mass transport theory over linear dynamics [8]. It follows that the trajectory of the covariances of the output is given by $\Sigma(t) = C\hat{\Sigma}(t)C^T$. Finally, let $\mu(t) = x(t)$ be the solution to (7), 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.

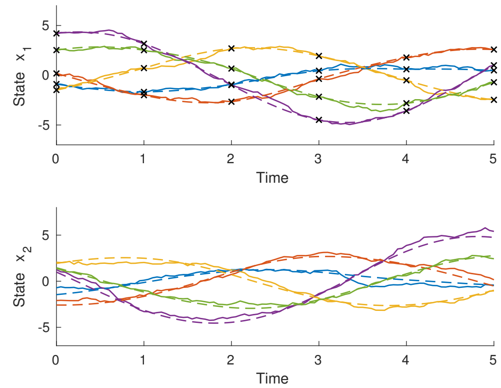


Fig. 1: Example with $N = 5$ systems to be tracked. Noise level: $\sigma = 0.5$. Available measurement points (x). True system states (solid). Estimated system stares (dashed) Upper figure: State x_1. Lower figure: State x_2.

A. Tracking of particles

We illustrate the tracking of a series of systems with a given system dynamic. Consider the tracking of $N = 5$ systems with oscillatory dynamics

$$dx(t) = Ax(t)dt + \sigma dw(t)$$
$$y(t) = Cx(t)$$

where the state dynamics is given by

$$A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad C = \begin{pmatrix} 1 & 0 \end{pmatrix}.$$

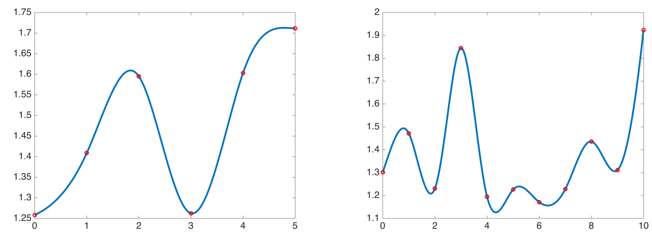
and where dw is normalized white Gaussian noise. We seek to recover the full state information of the systems based on only the unordered outputs, observed at the time point $t = 0, 1, \dots, 5$. For this example we use $\sigma = 0.5$. Figure 1 shows the reconstruction based on the optimization problem (6). 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

$$A = \begin{bmatrix} 0 & I \\ 0 & 0 \end{bmatrix}, B = \begin{bmatrix} 0 \\ I \end{bmatrix}, C = \begin{bmatrix} I & 0 \end{bmatrix},$$

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 (9). The results are depicted in Figure 2 for $T = 5, 10$. It can be observed that the interpolated curves are smooth. Similar results hold for high dimensional setting. Figure 3 depicts several Gaussian distributions with different means and covariances we want



(a) $T = 5$ (b) $T = 10$

Fig. 2: Interpolation of covariances

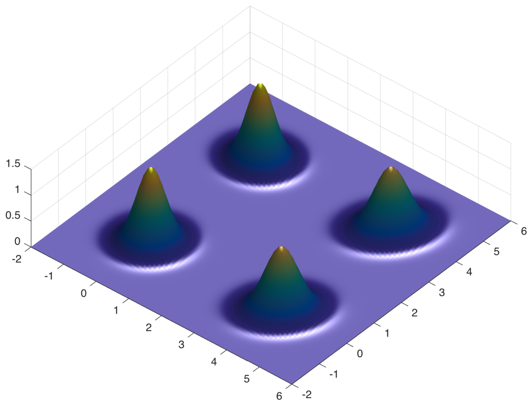


Fig. 3: Marginal distributions

to track. The tracking result is shown in Figure 4, 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.

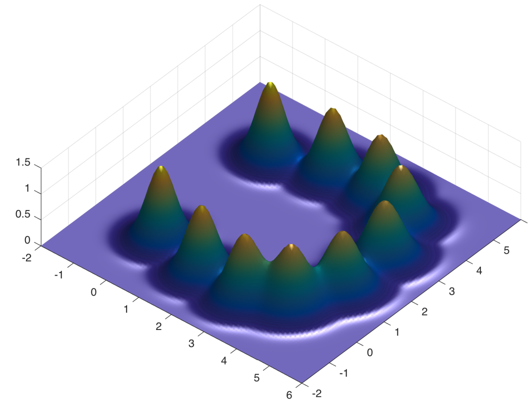


Fig. 4: Interpolation of covariances: $T = 3$

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] Y. Chen and J. Karlsson. State tracking of linear ensembles via optimal mass transport. *IEEE Control Systems Letters*, Published online April 13, 2018.

[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] 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.

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

[15] 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.

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

[17] 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.

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