Mixtures of ensembles: System separation and identification via optimal transport

Filip Elvander (Aalto University) and Isabel Haasler (Uppsala University)

Introduction

Many systems in control and biology applications involve large populations of agents, but only aggregated state observations are available. Examples include pedestrian flows in crowd dynamics [10, 15], swarms of animals [17], and population-level measurements in single-cell biology [2, 13, 7]. These populations are typically heterogeneous, comprising several ensembles governed by distinct dynamical models. Identifying the ensembles and associated dynamics from aggregate data may have a wide range of implications: in crowd dynamics, it enables forecasting and control of heterogeneous groups with different objectives [10, 12]; in biology, it allows for identifying subpopulations of cells with distinct developmental or disease trajectories from population-level measurements [1, 14], facilitating more targeted interventions. To the best of our knowledge, previous works on joint system identification of several systems consider only settings, where individual particle trajectories are available [4, 6, 9]. We propose a novel framework based on optimal transport that jointly infers ensemble assignments and identifies the dynamics of each ensemble from aggregate observations alone. We see a great potential for the proposed framework for problems in control theory, as well as related fields of research such as signal processing. For example, we have in a follow-up paper applied our framework to the problem of separating audio sources for real speech data [5].

Optimal transport has emerged as a powerful tool for modeling and controlling the evolution of distributions [3, 8], and has recently been applied to the inference of dynamics in homogeneous populations [11]. Consider two non-negative measures $\mu, \nu \in \mathcal{M}_+(X)$ over a space X with the same total mass. The optimal transport problem is to find the most efficient way to move the mass from μ to ν with respect to an underlying cost function $c : X \times X \to \mathbb{R}_+$, where $c(x, y)$ describes the cost for moving a unit mass from $x \in X$ to $y \in X$. The optimal transport between the given measures is described by a transport plan, which is a measure $m \in \mathcal{M}_+(X \times X)$, where $m(x, y)$ describes the amount of mass moved from $x \in X$ to $y \in X$. Thus, the optimal transport plan m is the solution to [16]

$$\begin{aligned} \underset{m\in\mathcal{M}_+(X\times X)}{\text{minimize}} \quad & \int_{X\times X} c(x,y)dm(x,y) \\ \text{subject to} \quad & \int_{A\times X} dm(x,y) = \int_A d\mu(x), \\ & \int_{X\times B} dm(x,y) = \int_B d\nu(x), \\ & \text{for all measurable sets } A,B\subset X. \end{aligned} \tag{1}$$

Note that the constraints in (1) impose that the transport plan m indeed transports the mass from μ to ν.

Problem formulation

We assume that each ensemble is described by a discrete-time dynamical system

$$x^{(t+1)} = \Phi_{\theta_k}(x^{(t)}), \quad \text{for } k = 1, \ldots, K, \tag{2}$$

For a pair of consecutive states $(x^{(t)}, x^{(t+1)})$ in the same ensemble, we define their transport cost as

$$c_\theta(x^{(t)}, x^{(t+1)}) = \left\| \Phi_\theta(x^{(t)}) - x^{(t+1)} \right\|_2^2.$$

In order to separate the populations, and identify their dynamics, we find a transport plan $m_k^{(t)}$ for each ensemble $k = 1, \ldots, K$ and each time step $t = 1, \ldots, T-1$, by solving the optimal transport problem

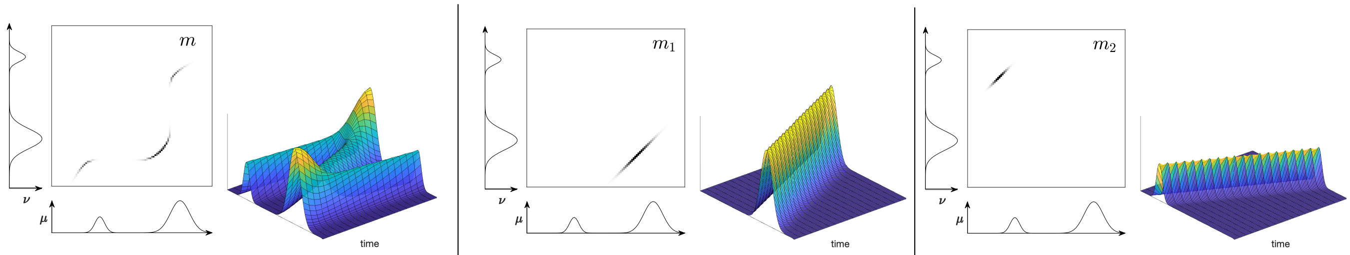
$$\begin{aligned} \underset{\substack{m_k^{(t)}\in\mathcal{M}_+(X\times X)\\ k=1,\ldots,K,t=1,\ldots,T-1\\ \theta_k\in\mathbb{R}^P,\ \mu_k^{(t)}\in\mathcal{M}_+(X)\\ k=1,\ldots,K,t=1,\ldots,T}}{\text{minimize}} \quad & \sum_{t=1}^{T-1}\sum_{k=1}^{K}\int_{X\times X} c_{\theta_k}(x,y)dm_k^{(t)}(x,y) \\ \text{subject to} \quad & \int_{A\times X} dm_k^{(t)}(x,y) = \int_A d\mu_k^{(t)}(x), \\ & \int_{X\times B} dm_k^{(t)}(x,y) = \int_B d\mu_k^{(t+1)}(x), \\ & \text{for all measurable sets } A,B\subset X,\ t=1,\ldots,T-1 \\ & \sum_{k=1}^{K}\mu_k^{(t)} = \mu^{(t)}, \quad t=1,\ldots,T, \end{aligned} \tag{3}$$

Example 1 *Consider two Gaussian mixtures μ and ν with two modes, where the means are switched. These modes can be understood as two subpopulations with different behaviors. Standard optimal transport does not use this information, as it minimizes the cost of moving the full distribution, see Figure 1a. We propose to search for two transport plans with parameter dependent cost assuming the dynamics $\Phi_{\theta_k}(x) = x+\theta_k$. Using these dynamics in (3) with $T = 2$ and $K = 2$, results in a unique minimizer with objective value 0, which is illustrated in Figure 1b. Each of the transport plans transports one of the modes.*

Proposition 1 *If the dynamics Φ_θ in (2) are linear in the parameter θ, then the non-convex problem (3) is bi-convex in the sets*

$$\left\{ \{m_k^{(t)}\}_{t=1}^{T-1}, \{\mu_k^{(t)}\}_{t=1}^{T} \right\}_{k=1}^{K} \quad \textit{and } \{\theta_k\}_{k=1}^K.$$

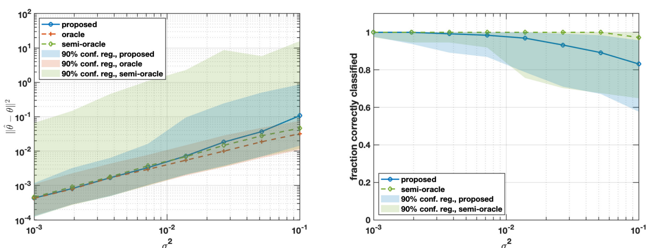
Thus, we solve the problem by a block coordinate descent method with the blocks as defined in the proposition. Under mild conditions, this method is guaranteed to converge to a local minimum.



(a) Classical optimal transport with cost function $c(x,y) = \|x-y\|_2^2$.

(b) Our method with underlying costs $c_{\theta_k}(x,y) = \|x-y+\theta_k\|_2^2$, for $k = 1, 2$, separates the transport into two transport plans m_1 and m_2, which each move one of the modes.

Figure 1: Comparison of classical and separated optimal transport between two Gaussian mixtures μ and ν. We show the given distributions and optimal the transport plans, where dark areas correspond to support in $\mathbb{R}^2$, as well as the corresponding evolutions of the distributions over time.



(a) Squared error for estimates of the ensemble dynamics parameters.

(b) The fraction of particles correctly grouped into their corresponding ensembles.

Figure 2: Simulation results of linear system identification of $K = 3$ ensembles over $T = 7$ time steps. The lines correspond to the median over 500 simulations, and the confidence regions cover 90% of the simulations.

Numerical experiments

We consider linear dynamics $x^{(t+1)} = Ax^{(t)} + b$ and generate aggregate observations of $K = 3$ ensembles with 10, 12, and 15 particles, respectively, over $T = 7$ time instances. In each simulation, the dynamics and initial conditions are drawn randomly, and Gaussian noise is added at every time step. Using our proposed method, we estimate both the dynamics and the ensemble assignments without assuming knowledge of the number of particles per ensemble. We run our algorithm from multiple random initializations and select the best solution. Figure 2a shows the estimation error for different noise levels, compared with two baselines: an oracle with access to true identities, and a semi-oracle that sees individual trajectories but not ensemble labels. Our method matches the oracle performance at low noise and maintains high classification accuracy even as noise increases (Figure 2b), outperforming the semi-oracle especially when clustering becomes unreliable.

References

[1] F. Buettner, K. N. Natarajan, F. P. Casale, V. Proserpio, A. Scialdone, F. J. Theis, S. A. Teichmann, J. C. Marioni, and O. Stegle. Computational analysis of cell-to-cell heterogeneity in single-cell rna-sequencing data reveals hidden subpopulations of cells. *Nature biotechnology*, 33(2):155–160, 2015.

[2] C. Bunne, S. G. Stark, G. Gut, J. S. Del Castillo, M. Levesque, K.-V. Lehmann, L. Pelkmans, A. Krause, and G. Rätsch. Learning single-cell perturbation responses using neural optimal transport. *Nature methods*, 20(11):1759–1768, 2023.

[3] Y. Chen, T. T. Georgiou, and M. Pavon. Optimal transport in systems and control. *Annual Review of Control, Robotics, and Autonomous Systems*, 4(1):89–113, 2021.

[4] Y. Chen and H. V. Poor. Learning mixtures of linear dynamical systems. In *International conference on machine learning*, pages 3507–3557. PMLR, 2022.

[5] L. Fabiani, S. J. Schlecht, I. Haasler, and F. Elvander. Joint spectrogram separation and TDOA estimation using optimal transport. *arXiv*, 2025.

[6] D. Fernex, B. R. Noack, and R. Semaan. Cluster-based network modeling—from snapshots to complex dynamical systems. *Science Advances*, 7(25):eabf5006, 2021.

[7] S. L. Goldman, M. MacKay, E. Afshinnekoo, A. M. Melnick, S. Wu, and C. E. Mason. The impact of heterogeneity on single-cell sequencing. *Frontiers in genetics*, 10:8, 2019.

[8] I. Haasler, J. Karlsson, and A. Ringh. Control and estimation of ensembles via structured optimal transport. *IEEE Control Systems Magazine*, 41(4):50–69, 2021.

[9] C. Hsu, M. Hardt, and M. Hardt. Linear dynamics: Clustering without identification. In *International Conference on Artificial Intelligence and Statistics*, pages 918–929. PMLR, 2020.

[10] P. Kachroo, S. J. Al-Nasur, S. A. Wadoo, and A. Shende. *Pedestrian dynamics: Feedback control of crowd evacuation*. Springer Science & Business Media, 2008.

[11] F. Lamoline, I. Haasler, J. Karlsson, J. Gonçalves, and A. Aalto. Gene regulatory network inference from single-cell data using optimal transport. *bioRxiv*, pages 2024–05, 2024.

[12] L. Scherrer, M. Tomko, P. Ranacher, and R. Weibel. Travelers or locals? Identifying meaningful sub-populations from human movement data in the absence of ground truth. *EPJ Data Science*, 7(1):1–21, 2018.

[13] G. Schiebinger, J. Shu, M. Tabaka, B. Cleary, V. Subramanian, A. Solomon, J. Gould, S. Liu, S. Lin, P. Berube, et al. Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell*, 176(4):928–943, 2019.

[14] D. Sun, X. Guan, A. E. Moran, L.-Y. Wu, D. Z. Qian, P. Schedin, M.-S. Dai, A. V. Danilov, J. J. Alumkal, A. C. Adey, et al. Identifying phenotype-associated subpopulations by integrating bulk and single-cell sequencing data. *Nature biotechnology*, 40(4):527–538, 2022.

[15] K. Szwaykowska, L. M.-y.-T. Romero, and I. B. Schwartz. Collective motions of heterogeneous swarms. *IEEE Transactions on Automation Science and Engineering*, 12(3):810–818, 2015.

[16] C. Villani. *Topics in optimal transportation*, volume 58. American Mathematical Soc., 2021.

[17] A. J. Ward, T. Schaerf, A. Burns, J. Lizier, E. Crosato, M. Prokopenko, and M. M. Webster. Cohesion, order and information flow in the collective motion of mixed-species shoals. *Royal Society open science*, 5(12):181132, 2018.