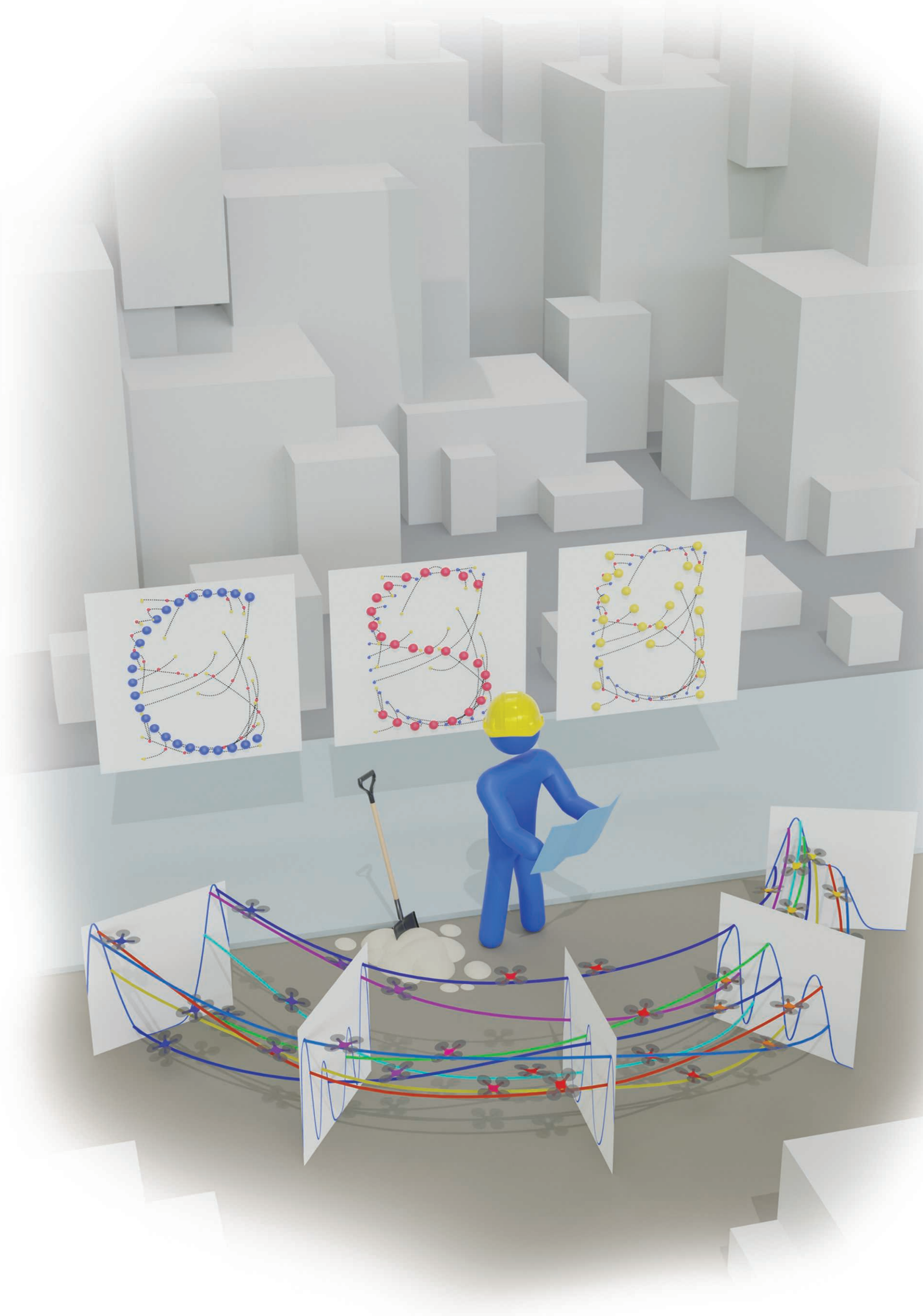
Optimal Transport for Applications in Control and Estimation

AN INTRODUCTION TO THE SPECIAL ISSUE

YONGXIN CHEN,
JOHAN KARLSSON,
and AXEL RINGH



In recent decades, there has been a rapid development of theory and methods for optimal transport, both within systems and control as well as in other areas such as signal processing, medical imaging, statistics, and machine learning. These advancements have developed optimal transport into an extensive framework with a large set of theoretical and computational tools that can be used to address problems in the areas of systems, control, and estimation. This special issue is organized to introduce optimal transport to a larger audience in the control community and summarize some of the recent progress in the field. The four articles in this issue present both theoretical and computational results, with a focus on the aspects relevant to the field of systems and control.

Digital Object Identifier 10.1109/MCS.2021.3076390
Date of current version: 19 July 2021

1066-033X/21©2021IEEE

This special issue is organized to introduce optimal transport to a larger audience in the control community and summarize some of the recent progress in the field.

The optimal transport problem is to determine an optimal way to transport one mass distribution to another, where the cost of moving a unit mass depends on the distance it is moved. This means deforming one distribution into another as cheaply as possible, where the price for moving mass depends both on the amount and distance moved. This minimal transport cost defines a distance between the two distributions [1].

Since this distance does not compare the distributions point by point (as do normal distances, such as L_p distances and the Kullback–Leibler divergence) but instead quantifies the required transport, it can be used for comparing distributions of different types (for example, comparing a density with a discrete set of agents). This property makes the distance a good candidate for quantifying uncertainty [2] and modeling deformations [3]. It also has appealing geometric properties, and when linking objects via geodesics of the optimal transport distance, there is a natural deformation between the objects that preserves "lumpiness." A graphical illustration of this is given in Figure 1, which compares a geodesic of the optimal transport distance with a geodesic of the L_2 distance.

Introducing geodesics can be seen as a first step in extending the traditional static optimal transport formulation (comparing two given distributions) to include dynamics. Further progress in this direction includes the reformulation of the optimal transport problem as a fluid dynamics problem [4], which can be interpreted as an optimal control problem for a density of particles or agents with simple dynamics. This interpretation can further be used to incorporate generalized underlying dynamics [5] and discrete dynamics over graphs [6]. All of these properties make the framework suitable for ensemble control and the control of densities as well as for estimation problems integrating data from a variety of sources or tracking moving objects [7]. This often results in optimization problems where the objective function contains one or more optimal transport costs [8]. Thus, it is also important to develop methods for solving such problems.

The optimal transport framework has many desirable properties, but it is often computationally challenging. The original formulation by Monge is a nonconvex optimization problem, while the later formulation by Kantorovich leads to a large-scale optimization problem that is intractable to solve with standard methods even for modest size problems. To address this computational problem, a recently popular approach is to add an entropic barrier term. The resulting optimization problem can then be solved using the so-called Sinkhorn iterations [9] that allow for computing an approximate solution to large transportation problems. This has opened up the field for new applications where no computationally feasible method previously exists.

Interestingly, the entropic regularized optimal transport coincides with another fascinating subject in physics started by Schrödinger, known as the Schrödinger bridge problem. This physics perspective brings tremendous insight into understanding the effects of entropic regularization in optimal transport, especially from a dynamical system perspective.

The following section provides a brief overview of the optimal transport problem. Specifically, the most common formulations of the problem are presented, with a focus on some of the properties that have made it such an interesting and useful tool in many areas (including control and estimation). The four articles in this issue are then summarized. Finally, a short exposition of the history related to the optimal transport problem is given in "A Brief History of Optimal Transport."

AN INTRODUCTION TO OPTIMAL TRANSPORT

Let $\mu_0, \mu_1 \in \mathcal{M}_+(X)$ be two nonnegative distributions defined on the state space $X \subset \mathbf{R}^d$ with the same total mass, that is, $\int_X \mu_0(x)dx = \int_X \mu_1(x)dx$. The optimal transport problem defines a notion of distance between the two distributions based on how costly it is to transform μ_0 into μ_1 by moving mass, where the cost of moving a unit mass is specified by a cost function $c(x_0, x_1)$. However, there are several different ways to mathematically formulate the problem of optimally moving the mass distributed as μ_0 into the distribution μ_1.

The first (the so-called Monge formulation) is based on the assumption that the transport plan is specified by a function $\phi: X \to X$, which means that the mass at point x is moved to the point $\phi(x)$. Monge's formulation of the optimal transport problem is then given by

$$\inf_{\phi} \int_X c(x, \phi(x)) \mu_0(x) dx, \tag{1a}$$

$$\text{subject to} \int_{x \in A} \mu_1(x) dx = \int_{\phi(x) \in A} \mu_0(x) dx \quad \text{for all measurable } A \subset X, \tag{1b}$$

where the total transportation cost (1a) for a transport plan ϕ is obtained by integrating the product of the mass moved [that is, $\mu_0(x)$] and the cost of moving the mass between

Introducing geodesics can be seen as a first step in extending the traditional static optimal transport formulation (comparing two given distributions) to include dynamics.

two points [that is, $c(x, \phi(x))$]. The transport plan ϕ must be such that μ_0 is transported to μ_1, which is ensured by the constraint (1b). Specifically, (1b) means that for any subsets $A \subset X$, the total mass of μ_1 in the set A is the same as the mass of μ_0 in the set that ϕ transport into A.

Although the Monge formulation (1) is an intuitive and straightforward definition, it has several shortcomings. For example, even if the marginals have the same total mass, there might not be a function ϕ that maps μ_0 to μ_1. Specifically, if μ_0 contains a point mass (that is, a Dirac delta function), then ϕ cannot split the mass into several points and therefore, (1b) is infeasible if μ_1 is a density. Further, the problem is nonconvex, and it is not symmetric in its arguments.

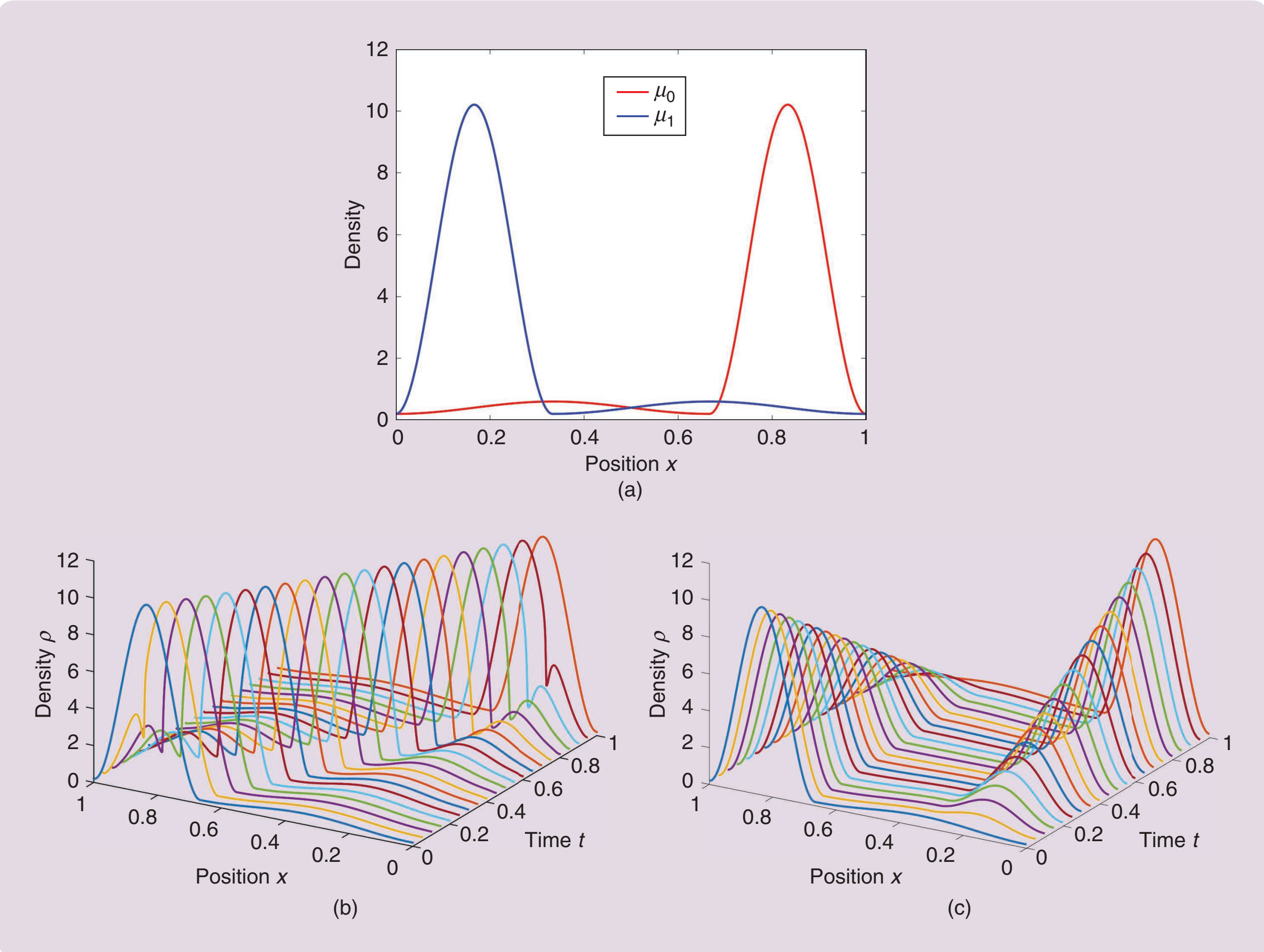


FIGURE 1 A numerical example illustrating the geodesics connecting the two distributions μ_0 and μ_1, which are shown in (a). A geodesic $\rho(t)$ can be understood as an optimal interpolation between the two distributions and corresponds to the optimal movement of the mass as a function of time for transforming the distribution μ_0 into the distribution μ_1. Different ways of measuring the distance between the two mass distributions, therefore, give rise to different geodesics $\rho(t)$. When the distance used is the optimal transport distance with underlying cost $c(x_0, x_1) = \| x_0 - x_1 \|_2^2$, the geodesic is the one shown in (b). When the L_2 distance is used, the obtained geodesic is the one shown in (c). As illustrated, the optimal transport geodesic better preserves the modality of the distribution in the intermediate time steps (that is, the geodesic better preserves the "lumpiness").

A Brief History of Optimal Transport

The subject of efficiently (optimally) allocating and transporting resources is an age-old subject. However, the first mathematical formulation of the optimal transport problem is attributed to the French mathematician and engineer Gaspard Monge (1746–1818). Monge was studying how to optimally transport soil for the construction of forts and bridges. In 1781, he formulated the problem as "divide two equal volumes into infinitesimal particles and associate them one to another so that the sum of the path lengths multiplied by the volumes of the particles be minimum possible."

As partly explained in this article, Monge's formulation (1) is mathematically challenging, and it took until the 1930s and 1940s before the next big breakthrough came. More precisely, in 1939, the Russian mathematician and economist Leonid Kantorovich (1912–1986) formulated and analyzed the optimal transport problem (2). Kantorovich developed a framework based on duality, where the variables in the dual problem to (2) represent the prices for selling and buying at the different locations. Hence, he is regarded as one of the founders of linear programming. For these (and other) achievements, he was the winner of the Stalin Prize in 1949, the Lenin Prize in 1965, and the Nobel Memorial Prize in Economic Sciences in 1975.

What followed was a period of rapid development of the theory for optimal transport, where fundamental properties of the problem were investigated and discovered. Important results include the existence and regularity of optimal solutions, polar decomposition (which characterizes the solution to the optimal transport problem), the introduction and investigation of displacement convexity (which later led to fruitful connections between optimal transport and curvature bound for Riemannian manifolds), the Benamou–Brenier dynamic formulation (3) of optimal transport, and the gradient flow formulation induced by optimal transport. Some of these results—as well as the use of these methods—have resulted in prestigious prizes such as the Fields Medal (C. Villani in 2010 and A. Figalli in 2018).

Lately, the focus of optimal transport research has shifted to algorithms and applications. Specifically, over the last decade there has been a massive amount of applicational work of optimal transport in areas such as imaging, machine learning, control, signals, and systems. Meanwhile, many efficient algorithms have been developed that made various applications possible. The most well-known algorithm that is widely used to solve optimal transport problems is the Sinkhorn algorithm.

The following formulation of Kantorovich overcomes these issues. Instead of using a function ϕ to define the transport plan, consider a nonnegative distribution on the product space $\pi \in \mathcal{M}_+(X \times X)$ (called the transport plan or coupling plan), where $\pi(x_0, x_1)$ denotes the amount of mass transported between the points x_0 and x_1. This allows for the optimal transport problem to be formulated as the linear programming problem

$$T(\mu_0, \mu_1) := \inf_{\pi \in \mathcal{M}_+(X\times X)} \int_{X\times X} c(x_0, x_1)\pi(x_0, x_1)\,dx_0 dx_1, \tag{2a}$$

$$\text{subject to} \int_X \pi(x_0, x)\,dx = \mu_0(x_0), \quad \text{for } x_0 \in X, \tag{2b}$$

$$\int_X \pi(x, x_1)\,dx = \mu_1(x_1), \quad \text{for } x_1 \in X. \tag{2c}$$

Specifically, note that this formulation is symmetric in the marginals μ_0 and μ_1 and that it allows for splitting mass in the allocation process. More precisely, in (2), mass in $\mu_0(x_0)$ can be moved to multiple points in the marginal μ_1 by simply making $\pi(x_0, x)$ nonzero for multiple values of x. The latter also explains the constraints (2b) and (2c); the mass moved from one point in one of the marginals must be received in the other marginal.

The formulation (2) also has several other useful properties. For example, the problem is convex, there is always a feasible solution π to (2), and the minimum exists as long as the cost function c is lower semicontinuous. In fact, (2) can be seen as a relaxation of (1). Specifically, if (1) has an optimal solution, then the corresponding mass transport plan of (2) is also optimal, and the optimal costs are the same in both problems.

A discrete version of problem (2) (where the marginals μ_0 and μ_1 are finite-dimensional vectors, and the transport plan π is a matrix) can be interpreted as a well-known minimum-cost flow problem on a complete bipartite graph [10, Lemma 9.3]. This is illustrated in Figure 2. Methods for numerically solving (2) can be derived based on discretizing the problem and solving the resulting finite-dimensional linear program.

However, such an approach has the drawback that the problem scales unfavorably in the dimension of the state space X and in the number of discretization points. More precisely, if $X \subset \mathbf{R}^d$ and each dimension is discretized into N points, then the marginals will be of dimension N^d, and the discrete transport plan will be of dimension $N^d \times N^d = N^{2d}$. Hence, the resulting linear programming problem has N^{2d} variables. Even for modest numbers such as $d = 2$ and $N = 100$, this results in 10^8 variables.

A recent approach for addressing this problem, which has been very successful, is to perturb the linear program with an entropy term [11]. The optimal solution to this perturbed problem, which is still a convex optimization problem, is highly structured. Specifically, this structure reduces the number of variables from N^{2d} to $2N^d$. Moreover,

These metrics can be used for quantifying the distances between discrete and continuous distributions (for example, to quantify when a particle cloud converges to a continuous density as the number of particles goes to infinity).

the optimal values of these variables can be found via the Sinkhorn iterations [11]. The latter is a simple, efficient, iterative procedure for the diagonal scaling of a nonnegative matrix to have given marginals, and this solution method has also motivated further research that has revealed deeper connections between optimal transport and the Schrödinger bridge problem [12].

The cost function c is defined on the underlying space X, and thus the optimal transport distance inherits properties of this space. Specifically, if the cost is selected as $c(x_0, x_1) = \|x_0 - x_1\|^p$ with $p > 0$ and where $\|\cdot\|$ is any norm on X, then $T(\mu_0, \mu_1)^{\min(1,1/p)}$ is a Wasserstein metric on the space of nonnegative distributions on X with the same total mass. These metrics can be used for quantifying the distances between discrete and continuous distributions (for example, to quantify when a particle cloud converges to a continuous density as the number of particles goes to infinity). Another useful property is that it allows for a shortest path interpolation, which, in these metrics, often corresponds to the smooth transportation of mass (in particular, when $p = 2$). An example illustrating this is the one given in Figure 1.

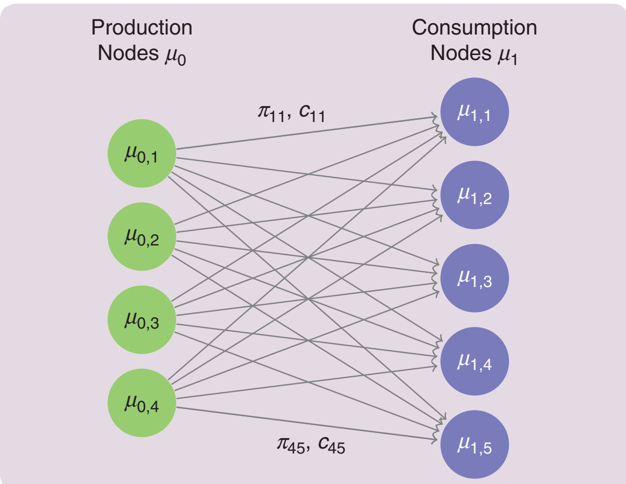


FIGURE 2 A small example of a discrete version of the optimal transport problem (2), illustrated as a minimum-cost flow problem on a complete bipartite graph. Here, the production in the four production nodes is represented by $\mu_0 \in \mathbf{R}_+^4$, and the consumption in the five consumption nodes is represented by $\mu_1 \in \mathbf{R}_+^5$. The flow between two nodes $\mu_{0,i}$ and $\mu_{1,j}$ is given by π_{ij}, and the transport plan $\pi = [\pi_{ij}]_{(i,j)=(1,1)}^{(4,5)}$ is feasible if there is flow balance (that is, if $\Sigma_{j=1}^5 \pi_{ij} = \mu_{0,i}$ for all i and $\Sigma_{i=1}^4 \pi_{ij} = \mu_{1,j}$ for all j). Moreover, the cost of transporting a unit from $\mu_{0,i}$ and $\mu_{1,j}$ is given by c_{ij}, and the total cost of transportation is thus $\Sigma_{(i,j)=(1,1)}^{(4,5)} c_{ij}\pi_{ij}$.

When the cost function is the squared Euclidean distance [that is, for $c(x_0, x_1) = \|x_0 - x_1\|_2^2$], the optimal transport problem can be created as an optimal control problem in fluid dynamics. This formulation (which is rather recent) was introduced in [4], where it was shown that when μ_0 and μ_1 are densities, both (1) and (2) are equivalent to

$$\inf_{\rho, v} \int_X \int_0^1 \rho(t,x) \|v(t,x)\|_2^2 dx dt, \tag{3a}$$

$$\text{subject to } \partial_t \rho + \nabla \cdot (\rho v) = 0, \tag{3b}$$

$$\rho(0,\cdot) = \mu_0, \quad \rho(1,\cdot) = \mu_1. \tag{3c}$$

Here, $\rho(t,\cdot)$ is the density of agents at time t, and $v(t,x)$ is the vector field representing the control action taken at a point x in the state space at time t. Since the amount of control needed in a point (t,x) is proportional to the density of agents, (3a) is the total cost of steering the density ρ between the given initial and final distributions in (3c). Moreover, $\nabla\cdot$ denotes the divergence of the vector field, and the constraint (3b) is a continuity equation that describes how the control affects the agent distribution and also guarantees that the latter has constant total mass.

The dynamic nature of the formulation in (3) makes it suitable for a range of applications in control and estimation. For instance, it is straightforward to rewrite it into a form familiar to those acquainted with stochastic control, namely as

$$\inf_v \mathbf{E}\left\{\int_0^1 \|v(t,x(t))\|_2^2 dt\right\}$$

$$\text{subject to } \dot{x}(t) = v(t, x(t)), \quad x(0) \sim \mu_0, \quad x(1) \sim \mu_1,$$

where the minimization is over all well-defined feedback control laws. This gives another interpretation of the illustration in Figure 1(b), namely as the expected evolution of a swarm of simple systems (integrators) when steered to transition from the initial to the final distribution using the combined, total minimum amount of energy. These formulations allow for generalizations to other dynamics, which will be addressed in several of the articles in this special issue.

CONTRIBUTIONS TO THIS SPECIAL ISSUE

This issue contains four articles focused on presenting optimal transport and its use in systems and control-related areas. The first article, by Amirhossein Taghvaei and Prashant G. Mehta, leverages optimal transport for understanding and constructing particle filtering algorithms. Specifically, they consider the feedback particle filter, where the Bayesian update step is done using a mean-field-type feedback control law instead of the conventional importance sampling and resampling. This updated step is understood using the notion of optimal coupling (a central concept in optimal transport theory) between prior and posterior distribution, and optimal transportation is then used to design the update rule.

The second article, by Isabel Haasler, Johan Karlsson, and Axel Ringh, uses structured multimarginal optimal transport to derive effective methods for optimal control and state estimation in multiagent systems. In fact, a duality result between control and estimation for multiagent systems is presented. The methods are designed by leveraging recent advances for both incorporating underlying dynamics into optimal transport problems and for efficiently solving the resulting large-scale optimization problems.

The third article, by Haomin Zhou, discusses optimal transport problems defined on networks and graphs. This is a relatively young area, and the article reviews some of these recent results and highlights some of the challenges of translating optimal transport onto networks and graphs. The practical implications of this effort are illustrated in an example where the concepts developed in the article are used to design a provably convergent algorithm for path exploration in an unknown environment in high-dimensional space.

The fourth article, by Yongxin Chen, Tryphon T. Georgiou, and Michele Pavon, is a survey of a recent line of research using optimal transport and its regularized version, the Schrödinger bridge theory, to control uncertainties in both continuous and discrete dynamics. For linear-quadratic Gaussian cases, this problem has been extensively studied under the name covariance control/steering. Applications of this emerging field include guidance and navigation in aerospace, active cooling of stochastic oscillators, robust transportation over networks, and many others.

AUTHOR INFORMATION

Yongxin Chen received the B.Sc. degree from Shanghai Jiao Tong University in 2011 and the Ph.D. degree from the University of Minnesota in 2016, both in mechanical engineering. He is an assistant professor in the School of Aerospace Engineering at Georgia Tech, Atlanta, Georgia, 30332, USA. Before joining Georgia Tech in 2018, he served on the faculty at Iowa State University (2017–2018). Prior to that, he was a postdoctoral researcher at Memorial Sloan Kettering Cancer Center in New York for one year. He received the George S. Axelby Best Article Award in 2017 for his joint work with Tryphon Georgiou and Michele Pavon. His current research focuses on the intersection of control, machine learning, optimization, and robotics.

Johan Karlsson received the M.Sc. degree in engineering physics from KTH Royal Institute of Technology in 2003 and the Ph.D. degree in optimization and systems theory from KTH in 2008. From 2009 to 2011, he was with Sirius International, Stockholm, and from 2011 to 2013, he was a postdoctoral associate in the Department of Computer and Electrical Engineering, University of Florida. He joined the Department of Mathematics, KTH, Stockholm, SE-100 44, Sweden, as an assistant professor in 2013 and has served as an associate professor since 2017. He has been the main organizer of several workshops for establishing collaborations between academia and industry. His current research interests include optimal transport, inverse problems and methods for large-scale optimization for applications in remote sensing, signal processing, and control theory.

Axel Ringh (eeringh@ust.hk) received the M.Sc. degree in engineering physics in 2014 and the Ph.D. degree in applied and computational mathematics in 2019, both from KTH Royal Institute of Technology. He is currently a Wallenberg postdoctoral researcher with the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China. His current research is within the areas of control theory and applied mathematics and includes optimal transport, analytic interpolation problems, moment problems, and methods for convex optimization.

REFERENCES

[1] C. Villani, *Topics in Optimal Transportation*. Providence, RI: American Mathematical Society, 2003.

[2] J. Karlsson and T. Georgiou, “Uncertainty bounds for spectral estimation,” *IEEE Trans. Autom. Control*, vol. 58, no. 7, pp. 1659–1673, 2013. doi: 10.1109/TAC.2013.2251775.

[3] T. ur Rehman, E. Haber, G. Pryor, J. Melonakos, and A. Tannenbaum, “3D nonrigid registration via optimal mass transport on the GPU,” *Med. Image Anal.*, vol. 13, no. 6, pp. 931–940, 2009. doi: 10.1016/j.media.2008.10.008.

[4] J.-D. Benamou and Y. Brenier, “A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem,” *Numerische Mathematik*, vol. 84, no. 3, pp. 375–393, 2000. doi: 10.1007/s002110050002.

[5] Y. Chen, T. Georgiou, and M. Pavon, “Optimal transport over a linear dynamical system,” *IEEE Trans. Autom. Control*, vol. 62, no. 5, pp. 2137–2152, 2017. doi: 10.1109/TAC.2016.2602103.

[6] S.-N. Chow, W. Huang, Y. Li, and H. Zhou, “Fokker–Planck equations for a free energy functional or Markov process on a graph,” *Arch. Ration. Mech. Anal.*, vol. 203, no. 3, pp. 969–1008, 2012. doi: 10.1007/s00205-011-0471-6.

[7] A. Taghvaei and P. Mehta, “An optimal transport formulation of the ensemble Kalman filter,” *IEEE Trans. Autom. Control*, early access, Aug. 10, 2020. doi: 10.1109/TAC.2020.3015410.

[8] J. Karlsson and A. Ringh, “Generalized Sinkhorn iterations for regularizing inverse problems using optimal mass transport,” *SIAM J. Imaging Sci.*, vol. 10, no. 4, pp. 1935–1962, 2017. doi: 10.1137/17M111208X.

[9] G. Peyré and M. Cuturi, “Computational optimal transport,” *Found. Trends[textregistered] Mach. Learn.*, vol. 11, nos. 5–6, pp. 355–607, 2019. doi: 10.1561/2200000073.

[10] B. Korte and J. Vygen, *Combinatorial Optimization*, 6th ed. Berlin, Heidelberg: Springer-Verlag, 2012.

[11] M. Cuturi, “Sinkhorn distances: Lightspeed computation of optimal transport,” *Adv. Neural Inform. Process. Syst.*, 2013, pp. 2292–2300.

[12] Y. Chen, T. Georgiou, and M. Pavon, “On the relation between optimal transport and Schrödinger bridges: A stochastic control viewpoint,” *J. Optim. Theory Appl.*, vol. 169, no. 2, pp. 671–691, 2016. doi: 10.1007/s10957-015-0803-z.