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

A relaxed maximum entropy approach to robust network routing *

Yongxin Chen[1], Tryphon Georgiou[2], Michele Pavon[3] and Allen Tannenbaum[4]

Abstract— **We consider network routing under random link failures with a desired final distribution. We provide a mathematical formulation of a relaxed transport problem where the final distribution only needs to be close to the desired one. The problem is a maximum entropy problem for path distributions with an extra terminal cost. We show that the unique solution may be obtained solving a *generalized Schrödinger system*. An iterative algorithm to compute the solution is provided. It contracts the Hilbert metric with contraction ratio less than $1/2$ leading to extremely fast convergence.**

I. INTRODUCTION

A relief organization, operating in an area where a natural disaster has occurred or in a war zone, is facing the following problem. At the initial time $t = 0$, there is a distribution $\nu_0(x)$ of available relief goods in sites $x \in \mathcal{X}$. Using the available road network, the goods must reach certain other locations after N units of time according to a desired distribution $\nu_N(x)$. Since the feasibility of the various possible routes is uncertain, it is desirable that the goods spread as much as the road network allows before reaching the target nodes. A certain flexibility in the final distribution can be afforded, thereby only requiring that it is *close* rather than equal to $\nu_N(x)$. In this paper, building on our previous work [15], [16], which deal with the case of a fixed terminal distribution, we provide a precise mathematical formulation of the above relaxed problem. It is a maximum entropy problem for probability distributions on the feasible paths with a terminal cost. We study a relaxed version of the usual Schrödinger bridge problem without a hard constraint on the terminal marginal but with an extra terminal cost. The solution is obtained by solving iteratively a *generalized Schrödinger system*. Convergence of the algorithm in a natural projective metric is established. In [28], which is a sort of relaxation of [11], the problem of optimally steering a linear stochastic system with a Wasserstein distance terminal cost was studied. In [18] (see also [31]), a regularized transport problem with very general boundary costs was considered and solved through iterative *Schrödinger-Fortet-Demin-Stephan-Sinkhorn-like* algorithms [43], [44], [26], [22], [45]. Although our dynamic problem can be reduced to a static one of the form considered in [18] using a well-known decomposition of relative entropy [24], [27, (2),p.033301-4], employing a general prior measure *on the trajectories* has some advantages. Indeed, the static formulation solution does not yield immediate by-product information on the new transition probabilities and on what paths the optimal mass flow occurs and is therefore less suited for many network routing applications. Moreover, we want to allow for general prior measures not necessarily of the Boltzmann's type considered in the previous work. Finally, we prove convergence of the iterative algorithm in the Hilbert [8] rather than Thompson metric as it usually provides the best contraction ratio.

We model the network through a directed graph and seek to design the routing policy so that the distribution of the commodity at some prescribed time horizon is close to a desired one. The optimal feedback control suitably modifies a prior transition mechanism. We also attempt to implicitly obtain other desirable properties of the optimal policy by suitably choosing a prior measure in a maximum entropy problem for distributions on paths. Robustness with respect to network failures, namely spreading of the mass as much as the topology of the graph and the final distribution allow, is accomplished by employing as prior transition the *adjacency matrix* of the graph. For other notions of robustness concerning networks see e.g. [1], [5], [4], [21], [42]. In particular, in [4], [21], robustness has been defined through a fluctuation-dissipation relation involving the entropy rate. This latter notion captures relaxation of a process back to equilibrium after a perturbation and has been used to study both financial and biological networks [40], [41]. This paper is addressed to transportation and data networks problems and does not concern equilibrium or near equilibrium cases.

In the next section, we define the relaxed transport problem. In Section III, we state the main result reducing the problem to solving a generalized Schrödinger system. In Section IV, we outline an iterative algorithm to compute the solution, some extensions of the results and provide one numerical example.

*This research was supported by AFOSR grants (FA9550-15-1-0045 and FA9550-17-1-0435), ARO grant (W911NF-17-1-049), grants from the National Center for Research Resources (P41-RR-013218) and the National Institute of Biomedical Imaging and Bioengineering (P41-EB-015902), NCI grant (1U24CA18092401A1), NIA grant (R01 AG053991), a grant from the Breast Cancer Research Foundation and by the University of Padova Research Project CPDA 140897.

[1] Yongxin Chen is with the Department of Electrical and Computer Engineering, Iowa State University, IA 50011, Ames, Iowa, USA yongchen@iastate.edu

[2] Tryphon Georgiou is with the Department of Mechanical and Aerospace Engineering, University of California, Irvine, CA 92697, USA tryphon@uci.edu

[3] Michele Pavon is with the Dipartimento di Matematica "Tullio Levi-Civita", Università di Padova, 35121 Padova, Italy pavon@math.unipd.it

[4] Allen Tannenbaum is with the Departments of Computer Science and Applied Mathematics & Statistics, Stony Brook University, NY 11794, USA allen.tannenbaum@stonybrook.edu

II. Relaxed Schrödinger bridges

Consider a directed, strongly connected aperiodic graph $\mathbf{G} = (\mathcal{X}, \mathcal{E})$ with vertex set $\mathcal{X} = \{1, 2, \ldots, n\}$ and edge set $\mathcal{E} \subseteq \mathcal{X} \times \mathcal{X}$. We let time vary in $\mathcal{T} = \{0, 1, \ldots, N\}$, and let $\mathcal{FP}_0^N \subseteq \mathcal{X}^{N+1}$ denote the family of length N, feasible paths $x = (x_0, \ldots, x_N)$, namely paths such that $(x_i, x_{i+1}) \in \mathcal{E}$ for $i = 0, 1, \ldots, N-1$.

We seek a probability distribution $\mathfrak{P}$ on $\mathcal{FP}_0^N$ with prescribed initial probability distribution $\nu_0(\cdot)$ and terminal distribution close to $\nu_N(\cdot)$, such that the resulting random evolution is closest to a "prior" measure $\mathfrak{M}$ on $\mathcal{FP}_0^N$ in a suitable sense. The prior law $\mathfrak{M}$ is induced by the Markovian evolution

$$\mu_{t+1}(x_{t+1}) = \sum_{x_t \in \mathcal{X}} \mu_t(x_t) m_{x_t x_{t+1}}(t) \tag{1}$$

with nonnegative distributions $\mu_t(\cdot)$ over $\mathcal{X}$, $t \in \mathcal{T}$, and weights $m_{ij}(t) \geq 0$ for all indices $i, j \in \mathcal{X}$ and all times. Moreover, to respect the topology of the graph, $m_{ij}(t) = 0$ for all t whenever $(i, j) \notin \mathcal{E}$. Often, but not always, the matrix

$$M(t) = [m_{ij}(t)]_{i,j=1}^n \tag{2}$$

does not depend on t. The rows of the transition matrix $M(t)$ do not necessarily sum up to one, so that the "total transported mass" is not necessarily preserved. This occurs, for instance, when $M(t)$ simply encodes the topological structure of the network with $m_{ij}(t)$ being zero or one, depending on whether a certain link exists at each time t. It is also possible to take into account the length of the paths leading to solutions which compromise between speading the mass and transporting on shorter paths, see [15], [16]. The evolution (1) together with the measure $\mu_0(\cdot)$, which we assume positive on $\mathcal{X}$, i.e.,

$$\mu_0(x) > 0 \text{ for all } x \in \mathcal{X}, \tag{3}$$

induces a measure $\mathfrak{M}$ on $\mathcal{FP}_0^N$ as follows. It assigns to a path $x = (x_0, x_1, \ldots, x_N) \in \mathcal{FP}_0^N$ the value

$$\mathfrak{M}(x_0, x_1, \ldots, x_N) = \mu_0(x_0) m_{x_0 x_1}(0) \cdots m_{x_{N-1} x_N}(N-1), \tag{4}$$

and gives rise to a flow of *one-time marginals*

$$\mu_t(x_t) = \sum_{x_{\ell \neq t}} \mathfrak{M}(x_0, x_1, \ldots, x_N), \quad t \in \mathcal{T}.$$

We seek a distribution which is closest to the prior $\mathfrak{M}$ in *relative entropy* where, for P and Q measures on $\mathcal{X}^{N+1}$, the relative entropy (divergence, Kullback-Leibler index) $\mathbb{D}(P\|Q)$ is

$$\mathbb{D}(P\|Q) := \begin{cases} \sum_{x \in \mathcal{X}^{N+1}} P(x) \log \frac{P(x)}{Q(x)}, & Supp(P) \subseteq Supp(Q), \\ +\infty, & Supp(P) \not\subseteq Supp(Q), \end{cases}$$

Here, by definition, $0 \cdot \log 0 = 0$. Naturally, while the value of $\mathbb{D}(P\|Q)$ may turn out negative due to miss-match of scaling (in case $Q = \mathfrak{M}$ is not a probability measure), the relative entropy is always jointly convex. Moreover,

$$\mathbb{D}(P\|Q) - \sum_{x \in \mathcal{X}^{N+1}} P(x) + \sum_{x \in \mathcal{X}^{N+1}} Q(x) \geq 0.$$

Since for probability distributions we have

$$\sum_{x \in \mathcal{X}^{N+1}} P(x) = 1,$$

minimizing the nonnegative quantity $\mathbb{D}(P\|Q) - \sum_x P(x) + \sum_x Q(x)$ over a family of probability distributions P, even when the prior Q has a different total mass, is equivalent to minimizing over the same set $\mathbb{D}(P\|Q)$. We are now ready to formalize the problem. Let ν_0 and ν_N be two probability distributions on $\mathcal{X}$ and let $\mathcal{P}(\nu_0)$ be the family of all Markovian probability distributions on $\mathcal{X}^{N+1}$ of the form (4) with initial marginal ν_0. Rather than imposing the desired final marginal ν_N as in the standard Schrödinger bridge problem, we consider the following "relaxed problem":

Problem 1:

$$\text{minimize } J(P) := \mathbb{D}(P\|\mathfrak{M}) + \mathbb{D}(p_N\|\nu_N) \tag{5a}$$

$$\text{over } \{P \in \mathcal{P}(\nu_0)\}. \tag{5b}$$

Clearly, we can restrict the minimization to distributions in $\mathcal{P}_S(\nu_0)$, namely distributions in $\mathcal{P}(\nu_0)$ such that

$$Supp(p_N) \subseteq Supp(\nu_N). \tag{6}$$

III. Main result

We have the following characterization of the solution.

Theorem 1: Assume that the matrix

$$G := M(N-1)M(N-2) \cdots M(1)M(0) = (g_{ij}) \tag{7}$$

has all positive elements g_{ij}. Suppose there exist two functions φ and $\hat{\varphi}$ mapping $\{0, 1, \ldots, N\} \times \mathcal{X}$ into the nonnegative reals and satisfying the *generalized Schrödinger system*

$$\varphi(t, i) = \sum_j m_{ij}(t) \varphi(t+1, j), \ 0 \leq t \leq N-1, \tag{8a}$$

$$\hat{\varphi}(t+1, j) = \sum_i m_{ij}(t) \hat{\varphi}(t, i), \ 0 \leq t \leq N-1, \tag{8b}$$

$$\varphi(0, i) \hat{\varphi}(0, i) = \nu_0(i), \tag{8c}$$

$$\varphi(N, j)^2 \hat{\varphi}(N, j) = \nu_N(j). \tag{8d}$$

For $0 \leq t \leq N-1$ and $(i, j) \in \mathcal{X} \times \mathcal{X}$, we define

$$\pi_{ij}^*(t) := m_{ij}(t) \frac{\varphi(t+1, j)}{\varphi(t, i)}. \tag{9}$$

which constitute a family of *bona fide* transition probabilities. Then, the solution $\mathfrak{P}^*$ to Problem 1 is unique and given by the Markovian distribution

$$\mathfrak{P}^*(x_0, \ldots, x_N) = \nu_0(x_0) \pi_{x_0 x_1}^*(0) \cdots \pi_{x_{N-1} x_N}^*(N-1). \tag{10}$$

The proof can be found in [17, Section III]. Existence and uniqueness for system (8) is established in [17, Section V] by resorting to Hilbert's projective metric.

IV. An algorithm contracting Hilbert's metric and some extensions

Let $\mathbf{1}^\dagger = (1, 1, \ldots, 1)$. System (8) suggests the following iterative algorithm:

a. Set $x = x(0) = \mathbf{1}$;
b. Set $x_{\text{next}} = \mathcal{C}(x)$;
c. Iterate until you reach a fixed point $\bar{x} = \mathcal{C}(\bar{x})$ (*stopping criterion:* $|\bar{x} - \mathcal{C}\bar{x}| < 10^{-4}$);
d. Set $\hat{\varphi}(N) = \bar{x}$;
e. Use
$$\varphi(N, x_N) = \sqrt{\frac{\nu_N(x_N)}{\hat{\varphi}(N, x_N)}} \tag{11}$$
to compute $\varphi(N)$ and then (8a) to compute $\varphi(t)$ for $t = N-1, N-2, \ldots, 1, 0$;
f. Compute the optimal transition probabilities $\pi_{ij}^*(t)$ according to (9);
g. The solution to Problem 1 is the time inhomogeneous Markovian distribution (10) with initial marginal ν_0 and transition probabilities $\pi_{ij}^*(t)$.

The assumption that the elements g_{ij} of the matrix $G = M(N-1)M(N-2)\cdots M(1)M(0)$ be all positive can be relaxed, see [17, Section VI].

Our analysis and algorithm can be generalized to the cost function
$$\mathbb{D}(P\|\mathfrak{M}) + \eta\mathbb{D}(p_N\|\nu_N) \tag{12}$$
for any $\eta \geq 0$. In this case, we only need change (8d) in the Schrödinger system to
$$\varphi(N, j)^{\frac{\eta+1}{\eta}}\hat{\varphi}(N, j) = \nu_N(j)$$
and (11) to
$$\varphi(N, x_N) = \left(\frac{\nu_N(x_N)}{\hat{\varphi}(N, x_N)}\right)^{\frac{\eta}{\eta+1}}$$
in the algorithm. The convergence rate is strictly upper bounded by $\frac{\eta}{\eta+1}$. The parameter η measures the significance of the penalty term $\mathbb{D}(p_N\|\nu_N)$. When η goes to infinity, we recover the traditional Schrödinger bridge. The upper bound is 1 in this case. On the other hand, when $\eta = 0$, the solution is trivial. It is the Markov process with kernel $M(t)$ (assuming that all $M(t)$ are stochastic matrices) and initial distribution ν_0.

Example 1: Consider the graph in Figure 1. We seek to transport masses from initial distribution $\nu_1 = \delta_1$ to target distribution $\nu_N = 1/2\delta_6 + 1/2\delta_9$. The step N is set to be 3 or 4. When $N = 3$, the evolution of mass distribution by solving Problem 1 is given by
$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.5865 & 0.2067 & 0.2067 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.3798 & 0 & 0.2067 & 0.4135 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.3798 & 0 & 0 & 0.6202 \end{bmatrix},$$

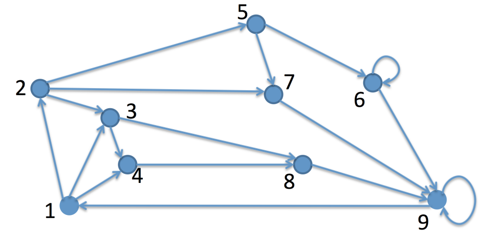

Fig. 1: transport graph

where the four rows of the matrix show the mass distribution at time step $t = 0, 1, 2, 3$ respectively. The prior law M is taken to be the Rulle Bowen random walk [15]. The mass spreads out before reaching nodes 6 and 9. Due to the soft terminal constraint, the terminal distribution is not equal to ν_N. When we allow for more steps $N = 4$, the mass spreads even more before reassembling at nodes $6, 9$, as shown below,
$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.6941 & 0.2040 & 0.1020 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.1020 & 0.1020 & 0.4901 & 0 & 0.1020 & 0.2040 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.3881 & 0.1020 & 0.2040 & 0.3059 \\ 0 & 0 & 0 & 0 & 0 & 0.2862 & 0 & 0 & 0.7138 \end{bmatrix}.$$
The terminal distribution is again not equal to ν_N. However, if we increase the penalty on $\mathbb{D}(p_N\|\nu_N)$, then the difference between p_N and ν_N becomes smaller, as can be seen below, which is the distribution evolution when $\eta = 10$ in (12)
$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.7679 & 0.1547 & 0.0774 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.0774 & 0.0774 & 0.6132 & 0 & 0.0774 & 0.1547 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.5359 & 0.0774 & 0.1547 & 0.2321 \\ 0 & 0 & 0 & 0 & 0 & 0.4585 & 0 & 0 & 0.5415 \end{bmatrix}.$$

V. Final comments

Since the work of Mikami, Thieullen, Leonard, Cuturi [36], [37], [38], [32], [33], [20], a large number of papers have appeared where Schrödinger bridge problems are viewed as regularization of the important Optimal Mass Transport (OMT) problem, see e.g., [7], [12], [13], [14], [34], [2], [18]. This is, of course, interesting and extremely effective as OMT is computationally challenging [3], [6]. Nevertheless, one should not forget that Schrödinger bridge problems have at least two other important motivations: The first is Schrödinger's original "hot gas experiment" model, namely *large deviations of the empirical distribution on paths* [24]. The second is a *maximum entropy principle in statistical inference*, namely choosing the a posterior distribution so as to make the fewest number of assumptions about what is beyond the available information. This inference method has been noticeably developed over the years by Jaynes, Burg, Dempster and Csiszár [29], [30], [9], [10], [23], [19]. It is this last concept which largely inspired the original approach taken in this paper and in [15], [16] although connections to OMT were made there.

References

[1] R. Albert, H. Jeong, and A.-L. Barabási, Error and attack tolerance of complex networks, *Nature*, **406**, pp. 378–382, 2000.

[2] J. Altschuler, J. Weed and P. Rigollet, Near-linear time approximation algorithms for optimal transport via Sinkhorn iteration. ArXiv e-prints, arXiv:1705.09634, 2017.

[3] S. Angenent, S. Haker and A. Tannenbaum, Minimizing flows for the Monge–Kantorovich problem, *SIAM J. Math. Anal.*, **35** (1), 61-97 (2003).

[4] L. Arnold, V. M. Gundlach and L. Demetrius, Evolutionary formalism for products of positive random matrices, *The Annals of Applied Probability*, **4**, 3, pp. 859–901, 1994.

[5] A.-L. Barabàsi, *Network Science*, 2014.

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

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

[8] G Birkhoff, Extensions of Jentzch's theorem, *Trans. Amer. Math. Soc.*, **85**, 219-227, 1957.

[9] J. P. Burg, Maximum entropy spectral analysis, in *Proc.37th Meet.Society of Exploration Geophysicists,* 1967. Reprinted in Modern Spectrum Analysis, D. G. Childers, Ed. New York: IEEE Press, 1978. pp. 34-41.

[10] J. Burg, D. Luenberger, and D. Wenger, Estimation of Structured Covariance Matrices, *Proceedings of the IEEE*, **70**, pp. 963–974, 1982.

[11] Y. Chen, T.T. Georgiou and M. Pavon, Optimal steering of a linear stochastic system to a final probability distribution, Part I, *IEEE Trans. Aut. Control*, **61**, Issue 5, 1158-1169, 2016.

[12] Y. Chen, T.T. Georgiou and M. Pavon, "On the relation between optimal transport and Schrödinger bridges: A stochastic control viewpoint", *J. Optim. Theory and Applic.*, **169** (2), 671-691, 2016.

[13] Y. Chen, T.T. Georgiou and M. Pavon, Entropic and displacement interpolation: a computational approach using the Hilbert metric, *SIAM J. on Applied Mathematics*, **76** (6), 2375-2396, 2016.

[14] Y. Chen, T. Georgiou and M. Pavon, Optimal transport over a linear dynamical system, *IEEE Trans. Aut. Contr.*, **62**, 2137-2152, 2017.

[15] Y. Chen, T. Georgiou, M. Pavon and A. Tannenbaum, Robust transport over networks, *IEEE Trans. Aut. Control*, **62**, 4675-4682, 2017.

[16] Y. Chen, T.T. Georgiou, M. Pavon and A. Tannenbaum, Efficient-robust routing for single commodity network flows, ArXiv e-prints, arXiv: 1701.07625v2, *IEEE Trans. Aut. Control*, to appear.

[17] Y. Chen, T.T. Georgiou, M. Pavon and A. Tannenbaum, Relaxed Schrödinger bridges and robust network routing, arXiv: 1801.07852.

[18] L. Chizat, G. Peyr, B. Schmitzer and F.-X. Vialard, Scaling algorithms for unbalanced optimal transport problems, ArXiv e-prints, arXiv:1607.05816v2, *Mathematics of Computation*, in press.

[19] I. Csiszár, "Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems," *The Annals of Statistics*, **19(4)**: 2032-2066, 1991.

[20] M.Cuturi, "Sinkhorn distances: lightspeed computation of pptimal transport," *Adv. in Neur. Inf. Proc. Systems*, 2292-2300, 2013.

[21] L. Demetrius and T. Manke, Robustness and network evolution: an entropic principle, *Physica A: Statistical Mechanics and its Applications*, **346**, 3, pp. 682–696, 2005.

[22] W. E. Deming and F. F. Stephan, On a least squares adjustment of a sampled frequency table when the expected marginal totals are known, *Ann. Math. Statist.*, **11**, pp. 427-444, 1940.

[23] A. P. Dempster, Covariance selection, *Biometrics*, **28**,157-175, 1972.

[24] H. Föllmer, Random fields and diffusion processes, in: *Ècole d'Ètè de Probabilitès de Saint-Flour XV-XVII*, edited by P. L. Hennequin, Lecture Notes in Mathematics, Springer-Verlag, New York, 1988, vol.1362,102-203.

[25] F. Forni and R. Sepulchre, Differentially Positive Systems, *IEEE Trans. Aut. Control*, **61**, Issue: 2, 346 - 359, 2016.

[26] R. Fortet, Résolution d'un système d'equations de M. Schrödinger, *J. Math. Pure Appl.* IX (1940), 83-105.

[27] T. T. Georgiou and M. Pavon, Positive contraction mappings for classical and quantum Schrödinger systems, *J. Math. Phys.*, **56**, 033301 (2015).

[28] A. Halder, and E.D.B. Wendel, Finite Horizon Linear Quadratic Gaussian Density Regulator with Wasserstein Terminal Cost, *Proc. of the 2016 American Control Conference*, 7249-7254, 2016.

[29] E. T. Jaynes, Information Theory and Statistical Mechanics, *Physical Review Series II*, **106** (4): 620630, 1957. doi:10.1103/PhysRev.106.620. MR87305, and Information Theory and Statistical Mechanics II, *Physical Review Series II*, **108** (2): 171190, 1957. doi:10.1103/PhysRev.108.171. MR96414.

[30] E. T. Jaynes. On the rationale of maximum-entropy methods. *Proceedings of the IEEE*, 70(9):939–952, Sept. 1982.

[31] Johan Karlsson, and Axel Ringh, Generalized Sinkhorn iterations for regularizing inverse problems using optimal mass transport, *SIAM Journal on Imaging Sciences* 10(4) : 1935-1962, 2017

[32] C. Léonard, From the Schrödinger problem to the Monge-Kantorovich problem, *J. Funct. Anal.*, 2012, **262**, 1879-1920.

[33] C. Léonard, A survey of the Schroedinger problem and some of its connections with optimal transport, *Discrete Contin. Dyn. Syst. A*, 2014, **34** (4): 1533-1574.

[34] W. Li, P. Yin and S Osher, Computations of optimal transport distance with Fisher information regularization, ArXiv e-prints, arXiv:1704.04605.

[35] C. Liverani and M. P. Wojtkowski, Generalization of the Hilbert metric to the space of positive definite matrices, *Pacific J. Math.*, **166**, no. 2, 339-355, 1994.

[36] T. Mikami, Monge's problem with a quadratic cost by the zero-noise limit of h-path processes, *Probab. Theory Relat. Fields*, **129**, (2004), 245-260.

[37] T. Mikami and M. Thieullen, Duality theorem for the stochastic optimal control problem., *Stoch. Proc. Appl.*, 116, 1815?1835 (2006).

[38] T. Mikami and M. Thieullen, Optimal Transportation Problem by Stochastic Optimal Control, *SIAM Journal of Control and Optimization*, **47**, N. 3, 1127-1139 (2008).

[39] M. Pavon and F. Ticozzi, Discrete-time classical and quantum Markovian evolutions: Maximum entropy problems on path space, *J. Math. Phys.*, **51**, 042104-042125 (2010).

[40] R.Sandhu, T. Georgiou, E. Reznik, L. Zhu, I. Kolesov, Y. Senbabaoglu, and A. Tannenbaum1, "Graph curvature for differentiating cancer networks," *Scientific Reports (Nature)*, vol. 5, 12323; doi: 10.1038/srep12323 (2015).

[41] R. Sandhu, T. Georgiou, and A. Tannenbaum, "Ricci curvature: An economic indicator for market fragility and systemic risk," *Science Advances*, vol. 2, doi: 10.1126/sciadv.1501495, 2016.

[42] K. Savla, G. Como, and M. A. Dahleh, "Robust network routing under cascading failures," *IEEE Trans. on Network Science and Engineering*, vol. 1, no. 1, 53-66, 2014.

[43] E. Schrödinger, Über die Umkehrung der Naturgesetze, *Sitzungsberichte der Preuss Akad. Wissen. Berlin, Phys. Math. Klasse* (1931), 144-153.

[44] E. Schrödinger, Sur la théorie relativiste de l'électron et l'interpretation de la mécanique quantique, *Ann. Inst. H. Poincaré* **2**, 269, 1932.

[45] R. Sinkhorn, A relationship between arbitrary positive matrices and doubly stochastic matrices, *Ann. Math. Statist.*, **35** (1964), 876-879.