This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCNS.2019.2935623, IEEE Transactions on Control of Network Systems

Relaxed Schrödinger bridges and robust network routing

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

***Abstract*—We seek network routing towards a desired final distribution that can mediate possible random link failures. In other words, we seek a routing plan that utilizes alternative routes so as to be relatively robust to link failures. To this end, 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 cast as a maximum entropy problem for probability distributions on paths with an added terminal cost. The entropic regularizing penalty aims at distributing the choice of paths amongst possible alternatives. We prove that the unique solution may be obtained by solving a *generalized Schrödinger system* of equations. An iterative algorithm to compute the solution is provided. Each iteration of the algorithm contracts the distance (in the Hilbert metric) to the optimal solution by more than $1/2$, leading to extremely fast convergence.**

I. Introduction

Containing the 2017-18 Southern California wild fires has been a major challenge for CAL FIRE involving dispatching hundreds of fire engines and thousands of fire fighters including some provided by ten other states. Efficiently dispatching the fire engines over a long period of time (the Thomas fire, for instance, burned for more than one month) is a difficult task. The problem can be roughly described as follows: At the initial time $t = 0$ we have a certain distribution of fire engines in certain locations (nodes). Within at most N time units, so as to provide the crew shift, the engines must reach through the available road network the various fire locations (other nodes). The distribution must guarantee the minimum force necessary to fight each specific fire. Considering the difficulties and hazards involved in reaching their destination, it seems reasonable to require that the final distribution of the fire engines be *close* (rather than equal) to a desired one. Another specification of the routing plan is robustness with respect to link failures. This could be accomplished by dispatching engines on alternative routes even when they are assigned the same destination.

This project was supported by AFOSR (FA9550-17-1-0435), ARO (W911NF-17-1-049), NSF (1509387, 1665031, 1807664, 1839441 and 1901599), 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 School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA 30332, yongchen@gatech.edu

[2] Tryphon T. 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

In this paper, building on our previous work [17], [18], which dealt with the case of a fixed terminal distribution, we provide a mathematical formulation of the above problem. It is cast as a maximum entropy problem for probability distributions on (admissible) paths with a cost on the discrepancy between the terminal distribution and the target one. Thus, the formulation represents a relaxed version of the usual Schrödinger bridge problem, where the hard constraint on the terminal marginal distribution is replaced by a soft terminal cost on discrepancy at the terminal point from a target distribution. We show that the solution is obtained by solving iteratively a *generalized Schrödinger system* and we establish convergence of the algorithm in the natural projective metric (Hilbert metric).

More specifically, we model the network through a directed graph that reflects admissible paths, and we 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 consists of suitable modification of a *prior transition mechanism*. This follows the paradigm put forth in [17]. The essence of the "prior" is to provide a baseline for what paths are desirable and, thereby, bequeth desirable properties on optimal policy. For instance, the prior can be decided based on simulations or field experience from past attempts in implementing previous transport protocols. It should be underscored that the prior does not need to be a probability measure, and can be selected to lessen preference from transit through specific paths and nodes. Thereby, 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 or a suitably weighted version of it.

Our notion of robustness of the routing policy should not be confused with other notions of robustness concerning networks which have been put forward and studied, see e.g. [1], [5], [47], [3], [24], [49]. In particular, in [3], [24], 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. This paper is addressed to transportation and data networks problems and does not concern equilibrium or near equilibrium cases.

Prior related work includes [32] where the authors proposed a relaxation of the problem of optimally steering a linear stochastic system studied in [15], by introducing a a Wasserstein distance-penalty to a desired terminal marginal distribution. Also, in [19] (see also [36]), a regularized transport

2325-5870 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCNS.2019.2935623, IEEE Transactions on Control of Network Systems

problem with very general boundary costs is considered and solved through iterative *Schrödinger-Fortet-Demin-Stephan-Sinkhorn-like* algorithms [50], [51], [29], [25], [53]. Although our dynamic problem can be reduced to a static one of the form considered in [19] (see Section II), employing a general prior measure *on the trajectories* has distinct advantages. Indeed, the static formulation does not readily provide as a by-product information on the new transition probabilities nor does it suggest the paths where the optimal mass flows, and therefore it is less suited for typical network routing applications. Moreover, in this work, we allow for general prior measures that are not necessarily of the Boltzmann's type and considered in previous works. Finally, we prove convergence of the iterative algorithm in the Hilbert rather than Thompson metric as this typically provides the best contraction ratio [14, Theorem 3.4], [37].

The outline of the paper is as follows. In Section II we define the relaxed transport problem. In Section III, we state and prove the main result reducing the problem to solving a generalized Schrödinger system. In Section IV-A, we review some fundamental concepts and results concerning Hilbert's projective metric. In Section IV-B, we establish existence and uniqueness of the solution to the generalized Schrödinger system through a contraction mapping principle. Finally, in Section IV-C, we outline an iterative algorithm to compute the solution and discuss possible extensions of the results.

II. Relaxed Schrödinger bridges

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

We seek a probability distribution P on $\mathscr{F}\mathscr{P}_0^N$ 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 $\mathscr{F}\mathscr{P}_0^N$ in a suitable sense. The prior law $\mathfrak{M}$ is induced by the Markovian evolution

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

with nonnegative distributions $\mu_t(\cdot)$ over $\mathscr{X}$, $t \in \mathscr{T}$, and weights $m_{ij}(t) \geq 0$ for all indices $i, j \in \mathscr{X}$ and all times. Moreover, to respect the topology of the graph, $m_{ij}(t) = 0$ for all t whenever $(i, j) \notin \mathscr{E}$. 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" may not be necessarily preserved. Another instance is 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; in this case, $M(t)$ does not reflect a probabilistic transition mechanism but can still be utilized as a reference measure. For another choice of a reference measure, one can employ a weighted adjacency matrix which weighs in the length of the paths leading to a solution which compromises between spreading the mass and transporting on shorter paths see [17], [18].

The evolution (1) together with a measure $\mu_0(\cdot)$, which we assume positive on $\mathscr{X}$, i.e.,

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

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

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

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

$$\mu_t(x_t) = \sum_{x_0, \dots, x_{t-1}, x_{t+1}, \dots, x_N} \mathfrak{M}(x_0, x_1, \dots, x_N), \quad t \in \mathscr{T}.$$

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

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

Here, by definition, $0 \cdot \log 0 = 0$, and $Supp(P)$ denotes the subset in $\mathscr{X}^{N+1}$ where P takes positive values. Naturally, while the value of $\mathbb{D}(P\|Q)$ may turn out negative due to mismatch of scaling (in case $Q = \mathfrak{M}$ is not a probability measure), the relative entropy is always jointly convex. Moreover,

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

Since for probability distributions we have

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

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 formulate the problem. In this, we seek to identify a Markovian evolution with a given specified marginal distribution ν_0 at $t = 0$, that is close to a prior $\mathfrak{M}$ while at the same time having end-point marginal that is close to a target distribution. To this end, we denote by $\mathscr{P}(\nu_0)$ the family of all Markovian probability distributions on $\mathscr{X}^{N+1}$ with initial marginal ν_0. The problem we address can now be stated as follows:

Problem 1. *Let ν_0 and ν_N be two specified probability distributions on $\mathscr{X}$. Minimize*

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

over $P \in \mathscr{P}(\nu_0)$, where p_N denotes the marginal of P at $t = N$.

2325-5870 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCNS.2019.2935623, IEEE Transactions on Control of Network Systems

In other words, rather than imposing the desired final marginal ν_N as in the standard Schrödinger bridge problem, we seek a probability law P with specified marginal at $t=0$, which is close to $\mathfrak{M}$ in the relative entropy sense and, at the same time, the terminal marginal distribution of P is also close to the desired target ν_N, again in the relative entropy sense. Thus, we seek to minimize a weighted (here, for simplicity, with equal weights) sum of the two distances.

It is easy to see that in Problem 1 we can restrict attention to distributions in $\mathscr{P}_S(\nu_0)$, namely to distributions in $\mathscr{P}(\nu_0)$ for which

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

The connection between Problem 1, which is in essence "dynamic," to a corresponding static reformulation, such as those considered in [19], can be arrived at as follows. Let P and Q be two probability distributions on $\mathscr{X}^{N+1}$. For $x=(x_0,x_1,\dots,x_N)\in\mathscr{X}^{N+1}$, consider the multiplicative decomposition

$$P(x)=P_{x_0,x_N}(x)p_{0N}(x_0,x_N),$$

where

$$P_{\bar{x}_0,\bar{x}_N}(x)=P(x|x_0=\bar{x}_0,x_n=\bar{x}_N)$$

is the probability on paths conditioned at given values at the two ends and

$$p_{0N}(x_0,x_N):=\sum_{x_1,\dots,x_{N-1}}P(x_0,x_1,\dots,x_N).$$

Assuming without loss of generality that p_{0N} is everywhere positive on $\mathscr{X}\times\mathscr{X}$, and similarly for q_{0N} corresponding to Q, we obtain

$$\mathbb{D}(P\|Q)=\sum_{x_0x_N}p_{0N}(x_0,x_N)\log\frac{p_{0N}(x_0,x_N)}{q_{0N}(x_0,x_N)} \tag{7}$$
$$+\sum_{x\in\mathscr{X}^{N+1}}p_{0N}(x_0,x_N)P_{x_0,x_N}(x)\log\frac{P_{x_0,x_N}(x)}{Q_{x_0,x_N}(x)}.$$

This is the sum of two nonnegative quantities. The second becomes zero if and only if $P_{x_0,x_N}(x)=Q_{x_0,x_N}(x)$ for all $x\in\mathscr{X}^{N+1}$. Thus, the minimizer corresponds to the choice $P^*_{x_0,x_N}(x)=Q_{x_0,x_N}(x)$ where the conditional probabilities with specified values of paths at the two ends (i.e., the "bridges") are the same, and therefore Problem 1 reduces to

$$\min_{\substack{p_{0N}\ \text{such that}\\ \sum_{x_N}p_{0N}(\cdot,x_N)=\nu_0(\cdot)}}\ \mathbb{D}(p_{0N}\|m_{0N})+\mathbb{D}(p_N\|\nu_N). \tag{8}$$

Here, as before,

$$m_{0N}:=\sum_{x_1,\dots,x_{N-1}}\mathfrak{M}(x_0,x_1,\dots,x_N).$$

We point out that the prior measure mass only needs to be positive and does not need to be a probability measure, i.e., the total mass does not need to be one. We now discuss the original formulation (5) as motivated in the introduction.

III. Main result

The solution to Problem 1 is given in the following statement. It requires solving a two-point boundary value problem for functions φ and $\hat{\varphi}$, which is assumed in the statement of the theorem. The existence of solutions to this side-problem will be established later on.

Theorem 1. *Suppose there exist two functions φ and $\hat{\varphi}$ mapping $\{0,1,\dots,N\}\times\mathscr{X}$ into the nonnegative reals and satisfying the* ***generalized Schrödinger system*** *below:*

$$\varphi(t,i)=\sum_j m_{ij}(t)\varphi(t+1,j),\ 0\le t\le N-1, \tag{9a}$$
$$\hat{\varphi}(t+1,j)=\sum_i m_{ij}(t)\hat{\varphi}(t,i),\ 0\le t\le N-1, \tag{9b}$$
$$\varphi(0,i)\hat{\varphi}(0,i)=\nu_0(i), \tag{9c}$$
$$\varphi(N,j)^2\hat{\varphi}(N,j)=\nu_N(j). \tag{9d}$$

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

$$\pi^*_{ij}(t):=m_{ij}(t)\frac{\varphi(t+1,j)}{\varphi(t,i)}. \tag{10}$$

which constitutes a family of bona fide *transition probabilities. Then, the solution P^* to Problem 1 is unique and given by the Markovian distribution*

$$P^*(x_0,\dots,x_N)=\nu_0(x_0)\pi^*_{x_0x_1}(0)\cdots\pi^*_{x_{N-1}x_N}(N-1). \tag{11}$$

Proof. Let $\pi_{ij}(t)$ be the transition probabilities of the measure $P\in\mathscr{P}_S(\nu_0)$. Then, using the multiplicative decomposition (4) for both measures we get the representation

$$\mathbb{D}(P\|\mathfrak{M})=\mathbb{D}(\nu_0\|\mu_0)+\sum_{k=0}^{N-1}\sum_{x_k}\mathbb{D}(\pi_{x_k,\cdot}(k)\|m_{x_k,\cdot}(k))p_k(x_k). \tag{12}$$

Since ν_0 is fixed for all $P\in\mathscr{P}_S(\nu_0)$, $\mathbb{D}(\nu_0\|\mu_0)$ is independent of the choice of $P\in\mathscr{P}_S(\nu_0)$.

Let $\varphi(\cdot,\cdot)$ be *space-time harmonic* for the prior transition mechanism, namely let φ satisfy recursion (9a) on $0\le t\le N-1$. We observe now that Problem 1 is equivalent to minimizing over $\mathscr{P}_S(\nu_0)$ the following

$$J_{\text{modified}}(P):=\sum_{k=0}^{N-1}\sum_{x_k}\mathbb{D}\left(\pi_{x_k,\cdot}(k)\|m_{x_k,\cdot}(k)\frac{\varphi(k+1,\cdot)}{\varphi(k,x_k)}\right)p_k(x_k)$$
$$+\sum_{x_N}\log\left[\frac{p_N(x_N)\varphi(N,x_N)}{\nu_N(x_N)}\right]p_N(x_N). \tag{13}$$

The calculation that allows seeing this fact amounts to cancelling/contracting the terms in a "telescopic" expansion of the summation

$$\sum_{k=0}^{N-1}\sum_{x_k,x_{k+1}}\left(\pi_{x_kx_{k+1}}(k)\log\frac{\varphi(k+1,x_{k+1})}{\varphi(k,x_k)}\right)p_k(x_k).$$

In view of (9d), (13) reduces to

$$J_{\text{modified}}(P)=\sum_{k=0}^{N-1}\sum_{x_k}\mathbb{D}\left(\pi_{x_k,\cdot}(k)\|m_{x_k,\cdot}(k)\frac{\varphi(k+1,\cdot)}{\varphi(k,x_k)}\right)p_k(x_k)$$
$$+\mathbb{D}(p_N\|\varphi(N,\cdot)\hat{\varphi}(N,\cdot)).$$

2325-5870 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCNS.2019.2935623, IEEE Transactions on Control of Network Systems

We next note that

$$\pi^*_{ij}(t) := m_{ij}(t)\frac{\varphi(t+1,j)}{\varphi(t,i)} \tag{14}$$

constitute a family of transition probabilities. Indeed, $\pi^*_{ij}(t) \geq 0$ and, by (9a),

$$\sum_j \pi^*_{ij}(t) = \sum_j m_{ij}(t)\frac{\varphi(t+1,j)}{\varphi(t,i)} = \frac{\varphi(t,i)}{\varphi(t,i)} = 1.$$

Moreover, it can be shown [31] that $\varphi(N,\cdot)\hat{\varphi}(N,\cdot)$ is a probability vector. Utilizing the property that the relative entropy between probability vectors is nonnegative, we conclude that $J_{\text{modified}}(P)$ is bounded below by 0. In addition, the minimum 0 is achieved by π^* in (14) and $p_N = \varphi(N,\cdot)\hat{\varphi}(N,\cdot)$ provided this pair π^*, p_N is compatible. That is, starting from ν_0, the Markov chain with transition probability π^* will reach p_N at $t = N$. We next prove this is true. In fact, we show below that $p_t = \varphi(t,\cdot)\hat{\varphi}(t,\cdot)$ by induction.

Clearly, $p_0 = \nu_0 = \varphi(0,\cdot)\hat{\varphi}(0,\cdot)$ holds by assumption. Supposing $p_t = \varphi(t,\cdot)\hat{\varphi}(t,\cdot)$, then

$$\begin{aligned} p_{t+1}(j) &= \sum_{ij} \pi^*_{ij}(t)p_t(i) \\ &= \sum_{ij} m_{ij}(t)\frac{\varphi(t+1,j)}{\varphi(t,i)}\varphi(t,i)\hat{\varphi}(t,i) \\ &= \varphi(t+1,j)\hat{\varphi}(t+1,j), \end{aligned}$$

where we have used (9b) in the last equality. This completes our proof. □

Thus, as we have seen in the above proof, at each time $t = 0,1,\ldots,N$ the marginal p^*_t of the solution factors as

$$p^*_t(i) = \varphi(t,i)\hat{\varphi}(t,i). \tag{15}$$

This holds as well in the usual setting of Schrödinger bridges, in that the one-time marginals can be constructed as the product of space-time harmonic functions, see e.g., [41], [40]. Here, however, we wish to highlight the fact that the final condition (9d) for the Schrödinger system *is substantially different from the standard one*, see e.g. [18]. As a consequence, we get from (9d) that

$$\varphi(N,x_N) = \sqrt{\frac{\nu_N(x_N)}{\hat{\varphi}(N,x_N)}}. \tag{16}$$

Now let $\varphi(t)$ and $\hat{\varphi}(t)$ denote the *column vectors* with entries $\varphi(t,i)$ and $\hat{\varphi}(t,i)$, respectively, and $i \in \mathscr{X}$. In matrix form, (9a), (9b) and (14) read

$$\varphi(t) = M(t)\varphi(t+1), \quad \hat{\varphi}(t+1) = M(t)^T\hat{\varphi}(t), \tag{17a}$$

and

$$\Pi(t) = [\pi_{ij}(t)] = \operatorname{diag}(\varphi(t))^{-1}M(t)\operatorname{diag}(\varphi(t+1)). \tag{17b}$$

IV. Generalized Schrödinger system & a computational framework

The key technical step in the main result in Theorem 1 is the existence of solutions to the generalized Schrödinger system of equations. This system departs from the general form of the Schrödinger system in that one of the boundary conditions, in equation (9d) is quadratic in the space-time harmonic function ϕ. The purpose of the following subsections are to establish the existence of solutions to the Generalized Schrödinger system as well as an algorithm for computing the solution. The key ingredient of what follows is the Hilbert projection metric, and the fact that the relations that arise induce contractive maps in the Hilbert metric. This echoes similar constructions for the standard Schrödinger problem. For the benefit of the reader we first present in Section IV-A a brief exposition of the Hilbert metric and some of its key properties. This is followed by Section IV-B where build the stage for proving the existence of solutions to (9) and stated in Theorem (6). Finally in Section IV-C we summarize the steps of an algorithm to constructing solutions along with its application to an academic example.

A. Background: Hilbert's projective metric

This metric dates back to 1895 [33]. A crucial contractivity result that permits to establish existence of solutions of equations on cones (such as the Perron-Frobenius theorem) was proven by Garrett Birkhoff in 1957 [8]. Important extensions of Birkhoff's result to nonlinear maps were provided by Bushell [13], [14]. Various other applications of the Birkhoff-Bushell result have been developed such as to positive integral operators and to positive definite matrices [14], [39]. More recently, this geometry has proven useful in various problems concerning communication and computations over networks (see [55] and the work of Sepulchre and collaborators [52], [10], [4] on consensus in non-commutative spaces and metrics for spectral densities) and in statistical quantum theory [48]. A recent survey on the applications in analysis is [39]. The use of the Hilbert metric is crucial in the nonlinear Frobenius-Perron theory [38]. A considerable further extension of the Perron-Frobenius theory beyond linear positive systems and monotone systems has been recently proposed in [28].

Taking advantage of the Birkhoff-Bushell results on contractivity of linear and nonlinear maps on cones, we showed in [31] that the Schrödinger bridge for Markov chains and quantum channels can be efficiently obtained from the fixed-point of a map which is contractive in the *Hilbert metric*. This result extended [30] which deals with scaling of nonnegative matrices. In [16], it was shown that a similar approach can be taken in the context of diffusion processes leading to i) a new proof of a classical result on SBP and ii) providing an efficient computational scheme for both, SBP and OMT. This new computational approach can be effectively employed, for instance, in image interpolation.

2325-5870 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCNS.2019.2935623, IEEE Transactions on Control of Network Systems

In this section, following [14], we recall some basic concepts and results of this theory.

Let $\mathscr{S}$ be a real Banach space and let $\mathscr{K}$ be a closed solid cone in $\mathscr{S}$, i.e., $\mathscr{K}$ is closed with nonempty interior $\text{int}\mathscr{K}$ and is such that $\mathscr{K}+\mathscr{K}\subseteq\mathscr{K}$, $\mathscr{K}\cap-\mathscr{K}=\{0\}$ as well as $\lambda\mathscr{K}\subseteq\mathscr{K}$ for all $\lambda\geq 0$. Define the partial order

$$x\preceq y\Leftrightarrow y-x\in\mathscr{K},\quad x<y\Leftrightarrow y-x\in\text{int}\mathscr{K}$$

and for $x,y\in\mathscr{K}_0:=\mathscr{K}\backslash\{0\}$, define

$$\begin{aligned}M(x,y) &:= \inf\{\lambda\mid x\preceq\lambda y\}\\ m(x,y) &:= \sup\{\lambda\mid \lambda y\preceq x\}.\end{aligned}$$

Then, the Hilbert metric is defined on $\mathscr{K}_0$ by

$$d_H(x,y):=\log\left(\frac{M(x,y)}{m(x,y)}\right).$$

Strictly speaking, it is a *projective* metric since it is invariant under scaling by positive constants, i.e., $d_H(x,y)=d_H(\lambda x,\mu y)=d_H(x,y)$ for any $\lambda>0,\mu>0$ and $x,y\in\text{int}\mathscr{K}$. Thus, it is actually a distance between rays. If U denotes the unit sphere in $\mathscr{S}$, $(\text{int}\mathscr{K}\cap U,d_H)$ is a metric space.

Example 1. *Let $\mathscr{K}=\mathbb{R}^n_+=\{x\in\mathbb{R}^n:x_i\geq 0\}$ be the positive orthant of $\mathbb{R}^n$. Then, for $x,y\in\text{int}\mathbb{R}^n_+$, namely with all positive components,*

$$M(x,y)=\max_i\{x_i/y_i\},\quad m(x,y)=\min_i\{x_i/y_i\},$$

and

$$d_H(x,y)=\log\max\{\frac{x_iy_j}{y_ix_j}\}.$$

Another very important example for applications in many diverse areas of statistics, information theory, control, etc., is the cone of Hermitian, positive semidefinite matrices.

Example 2. *Let $\mathscr{S}=\{X=X^\dagger\in\mathbb{C}^{n\times n}\}$, where $\dagger$ denotes here transposition plus conjugation and, more generally, adjoint. Let $\mathscr{K}=\{X\in\mathscr{S}:X\geq 0\}$ be the positive semidefinite matrices. Then, for $X,Y\in\text{int}\mathscr{K}$, namely positive definite, we have*

$$d_H(X,Y)=\log\frac{\lambda_{\max}\left(XY^{-1}\right)}{\lambda_{\min}\left(XY^{-1}\right)}=\log\frac{\lambda_{\max}\left(Y^{-1/2}XY^{-1/2}\right)}{\lambda_{\min}\left(Y^{-1/2}XY^{-1/2}\right)}.$$

It is closely connected to the Riemannian (Fisher-information) metric

$$\begin{aligned}d_R(X,Y) &= \|\log\left(Y^{-1/2}XY^{-1/2}\right)\|_F\\ &= \sqrt{\sum_{i=1}^n[\log\lambda_i\left(Y^{-1/2}XY^{-1/2}\right)]^2}.\end{aligned}$$

A map $\mathscr{E}:\mathscr{K}\to\mathscr{K}$ is called *non-negative*. It is called *positive* if $\mathscr{E}:\text{int}\mathscr{K}\to\text{int}\mathscr{K}$. If $\mathscr{E}$ is positive and $\mathscr{E}(\lambda x)=\lambda^p\mathscr{E}(x)$ for all $x\in\text{int}\mathscr{K}$ and positive λ, $\mathscr{E}$ is called *positively homogeneous of degree p* in $\text{int}\mathscr{K}$. For a positive map $\mathscr{E}$, the *projective diameter* is befined by

$$\Delta(\mathscr{E}):=\sup\{d_H(\mathscr{E}(x),\mathscr{E}(y))\mid x,y\in\text{int}\mathscr{K}\}$$

and the *contraction ratio* by

$$k(\mathscr{E}):=\inf\{\lambda:\mid d_H(\mathscr{E}(x),\mathscr{E}(y))\leq\lambda d_H(x,y),\forall x,y\in\text{int}\mathscr{K}\}.$$

Finally, a map $\mathscr{E}:\mathscr{S}\to\mathscr{S}$ is called *monotone increasing* if $x\leq y$ implies $\mathscr{E}(x)\leq\mathscr{E}(y)$.

Theorem 2 ([14])**.** *Let $\mathscr{E}$ be a monotone increasing positive mapping which is positive homogeneous of degree p in $\text{int}\mathscr{K}$. Then, the contraction $k(\mathscr{E})$ does not exceed p. In particular, if $\mathscr{E}$ is a positive linear mapping, $k(\mathscr{E})\leq 1$.*

Theorem 3 ([8], [14])**.** *Let $\mathscr{E}$ be a positive linear map. Then*

$$k(\mathscr{E})=\tanh(\frac{1}{4}\Delta(\mathscr{E})).\tag{18}$$

Theorem 4 ([14])**.** *Let $\mathscr{E}$ be either*

a. *a monotone increasing positive mapping which is positive homogeneous of degree $p(0<p<1)$ in $\text{int}\mathscr{K}$, or*
b. *a positive linear mapping with finite projective diameter.*

Suppose the metric space $Y=(\text{int}\mathscr{K}\cap U,d_H)$ is complete. Then, in case (a) there exists a unique $x\in\text{int}\mathscr{K}$ such that $\mathscr{E}(x)=x$, in case (b) there exists a unique positive eigenvector of $\mathscr{E}$ in Y.

This result provides a far-reaching generalization of the celebrated Perron-Frobenius theorem [9]. Notice that in both Examples 1 and 2, the space $Y=(\text{int}\mathscr{K}\cap U,d_H)$ is indeed complete [14].

There are other metrics which are contracted by positive monotone maps. For instance, the closely related *Thompson metric* [54]

$$d_T(x,y)=\log\max\{M(x,y),m^{-1}(x,y)\}.$$

The Thompson metric is a *bona fide* metric on $\mathscr{K}$. It has been, for instance, employed in [43], [19], [4].

B. Solution to the generalized Schrödinger system

Let $G=M(0)M(1)\cdots M(N-2)M(N-1)=[g_{ij}]_{ij=1}^n$ and assume that all its elements are positive. Let us introduce the following maps on $\mathbb{R}^n_+$:

$$\mathscr{E}:\quad x\mapsto y\qquad\text{where } y_i=\sum_j g_{ij}x_j,\tag{19a}$$

$$\mathscr{E}^\dagger:\quad x\mapsto y\qquad\text{where } y_j=\sum_i g_{ij}x_i,\tag{19b}$$

$$\mathscr{D}_0:\quad x\mapsto y=\tfrac{v_0}{x}\quad\overset{\text{def}}{\Leftrightarrow} y_i=\frac{(v_0)_i}{x_i}\tag{19c}$$

$$\mathscr{D}_N:\quad x\mapsto y=\tfrac{v_N}{x}\quad\overset{\text{def}}{\Leftrightarrow} y_i=\frac{(v_N)_i}{x_i}\tag{19d}$$

2325-5870 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCNS.2019.2935623, IEEE Transactions on Control of Network Systems

i.e. division of vectors is carried out componentwise[1].

Lemma 1. *Consider the maps $\mathscr{E}$ and $\mathscr{E}^\dagger$. We have the following bounds on their contraction ratios:*

$$k(\mathscr{E}) = k(\mathscr{E}^\dagger) = \tanh(\frac{1}{4}\Delta(\mathscr{E})) < 1. \tag{20}$$

Proof. Observe that $\mathscr{E}$ is a positive *linear* map and its projective diameter is

$$\begin{aligned}\Delta(\mathscr{E}) &= \sup\{d_H(\mathscr{E}(u),\mathscr{E}(v)) \mid u,v \in \mathbb{R}^n_+\} \\ &= \sup\{\log\left(\frac{g_{ik}g_{j\ell}}{g_{i\ell}g_{jk}}\right) \mid 1 \le i,j,k,\ell \le n\}.\end{aligned}$$

To see the validity of the second line note that the supremum is attained when each u,v select a suitable row of G (e.g., u,v each having a single nonzero entry at position k and ℓ, respectively). Then set $x = \mathscr{E}(u)$ and $y = \mathscr{E}(v)$ and utilize the expression in Example 1.

The expression for $\Delta(\mathscr{E})$ is now seen to be finite, since all entries g_{ij}'s are positive. It now follows from Theorem 3 that its contraction ratio satisfies (20). A similar conclusion holds for the adjoint map $\mathscr{E}^\dagger$. □

Lemma 2.

$$k(\mathscr{D}_0) \le 1, \quad k(\mathscr{D}_N) \le 1$$

Proof. See [31, p.033301-10]. □

Lemma 3. *Let $\mathscr{R} : \mathbb{R}^n_+ \to \mathbb{R}^n_+$ be the map which associates to the vector x with components x_i to the vector with components $\sqrt{x_i}$, i.e.,*

$$\mathscr{R} : x \mapsto y = \sqrt{x}. \tag{21}$$

Then

$$k(\mathscr{R}) = 1/2. \tag{22}$$

Proof. Let $x,y \in \text{int}\mathbb{R}^n_+$. In view of Example 1 and using the properties of the square root,

$$d_H(\mathscr{R}(x),\mathscr{R}(y)) = \log\max\{\sqrt{(x_iy_j/y_ix_j)}\} = (1/2)d_H(x,y).$$

□

Theorem 5. *The composition*

$$\mathscr{C} := \mathscr{E}^\dagger \circ \mathscr{D}_0 \circ \mathscr{E} \circ \mathscr{R} \circ \mathscr{D}_N \tag{23}$$

contracts the Hilbert metric with contraction ratio $k(\mathscr{C}) < (1/2)$, namely

$$d_H(\mathscr{C}(x),\mathscr{C}(y)) < (1/2)d_H(x,y), \quad \forall x,y \in \text{int}\mathbb{R}^n_+.$$

Proof. The result follows at once from Lemmas 1, 2, 3. □

We now conclude with the following result that claims the existence of sought solutions to the Generalized Schrödinger system.

Theorem 6. *Assume that the elements of the matrix*

$$G = M(0)M(1)\cdots M(N-2)M(N-1) = [g_{ij}]_{ij=1}^n$$

are positive. Let ν_0 and ν_N be any two probability distributions on $\mathscr{X}$. Then, there exist a unique choice of the vectors $\varphi(0), \hat{\varphi}(N)$ with positive entries such that

$$\varphi(t,i) = \sum_j m_{ij}(t)\varphi(t+1,j),\ 0 \le t \le N-1 \tag{24a}$$

$$\hat{\varphi}(t+1,j) = \sum_i m_{ij}(t)\hat{\varphi}(t,i),\ 0 \le t \le N-1 \tag{24b}$$

$$\varphi(0,x_0)\hat{\varphi}(0,x_0) = \nu_0(x_0), \tag{24c}$$

$$\varphi(N,x_N)^2\hat{\varphi}(N,x_N) = \nu_N(x_N). \tag{24d}$$

Proof. Consider the iteration

$$(\hat{\varphi}(N,\cdot))_{\text{next}} = \mathscr{C}(\hat{\varphi}(N,\cdot)) \tag{25}$$

Notice that the componentwise divisions of $\mathscr{D}_0$ and $\mathscr{D}_N$ are well defined. Indeed, even when $\hat{\varphi}(0)$ ($\varphi(N)$) has zero entries, $\hat{\varphi}(N)$ ($\varphi(0)$) has all positive entries since the elements of G are all positive. Since $\mathscr{C}$ is strictly contractive in the Hilbert metric, the iteration (25) would converge to a ray that is invariant under $\mathscr{C}$. We next show that the iteration (25) in fact has a unique fixed point. Let $\hat{\phi}$ be any positive vector on the invariant ray, then

$$\mathscr{C}(\hat{\phi}) = \lambda\hat{\phi}$$

for some positive number λ. At the same time, by tracing all operations that make up $\mathscr{C}$, it is easy to see that

$$\mathscr{C}(\lambda^2\hat{\phi}) = \lambda\mathscr{C}(\hat{\phi}).$$

Combining these, let $x = \lambda^2\hat{\phi}$ and verify that

$$\mathscr{C}(x) = \lambda\mathscr{C}(\hat{\phi}) = \lambda^2\hat{\phi} = x. \tag{26}$$

On the other hand, for any $y = \alpha x$ with $\alpha \neq 1$, we get

$$\mathscr{C}(y) = \sqrt{\alpha}\mathscr{C}(x) = \sqrt{\alpha}x \neq y.$$

Hence, x is the unique fixed point of (25). Define

$$\begin{aligned}\hat{\varphi}(N) &= x \\ \varphi(N) &= \sqrt{\frac{\nu_N}{\hat{\varphi}(N)}} \\ \varphi(0) &= \mathscr{E}(\varphi(N)) \\ \hat{\varphi}(0) &= \frac{\nu_0}{\varphi(0)}\end{aligned}$$

and $\varphi(t), \hat{\varphi}(t)$ according to (24a)-(24b). Observe that these vectors are consistent with the Schrödinger system (24a-24d), and this completes the proof. □

[1]Our use of the adjoint for the map $\mathscr{E}$ is consistent with the standard notation in diffusion processes where the Fokker-Planck (forward) equation involves the adjoint of the *generator* appearing in the backward Kolmogorov equation.

2325-5870 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCNS.2019.2935623, IEEE Transactions on Control of Network Systems

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

We delineate the steps of the algorithm that results for the iterative steps suggested in (25) below. The sought result is to compute the transition probabilities $\pi^*_{ij}(t)$ in (10) via solving (24a). The transition probabilities represent the scheduling for the transportation problem that we have been considering. The summary of the algorithmic steps are followed by discussing a generalization that aims at weighing differently the terms of the optimization problem (Problem 1). We proceed with the basic algorithm that can be stated as follows.

Algorithm 1: Determine $\pi^*_{ij}(t)$ in (10)

Result: routing scheduling specified by $\pi^*_{ij}(t)$
Data:
matrices $M(t)$ in (2) and probability vectors ν_0, ν_N.
Initialization: $\mathscr{C}$ as defined in (23) where $\mathscr{D}_0, \mathscr{E}, \mathscr{R}, \mathscr{D}_N$ are as in (19) and (21).
Set stopping accuracy, e.g., eps $= 10^{-4}$.
Set $x = \mathbb{1} := (1,1,\dots,1)^\dagger$, and $x_{\text{next}} = \mathscr{C}(x)$.
while $|x - x_{\text{next}}| >$ eps **do**
  $x \leftarrow x_{\text{next}}$
  $x_{\text{next}} \leftarrow \mathscr{C}(x)$
end
$\hat{\varphi}(N) \leftarrow x_{\text{next}}$
$\varphi(N, x_N) \leftarrow \sqrt{\frac{\nu_N(x_N)}{\hat{\varphi}(N,x_N)}}$
$t \leftarrow N$
while $t > 0$ **do**
  $\varphi(t,i) = \sum_j m_{ij}(t)\varphi(t+1,j)$
  $t \leftarrow t-1$
end
while $t < N$ **do**
  $\pi^*_{ij}(t) := m_{ij}(t)\frac{\varphi(t+1,j)}{\varphi(t,i)}$
  $t \leftarrow t+1$
end

The assumption that the elements g_{ij} of the matrix $G = M(0)M(1)\cdots M(N-2)M(N-1)$ be all positive can be relaxed. For instance, if both ν_0 and ν_N are everywhere positive on $\mathscr{X}$, it suffices that G has at least one positive element in each row and column to guarantee that the componentwise divisions of $\mathscr{D}_0$ and $\mathscr{D}_N$ are well defined. In that case, Theorems 1 and 5 hold true and the algorithm of this section applies.

Our analysis and algorithm can be generalized to the cost function

$$\mathbb{D}(P\|\mathfrak{M}) + \eta\mathbb{D}(p_N\|\nu_N) \tag{27}$$

for any $\eta \geq 0$. In this case, we only need to change (9d) in the Schrödinger system to

$$\varphi(N,j)^{\frac{\eta+1}{\eta}}\hat{\varphi}(N,j) = \nu_N(j)$$

and the relation $\varphi(N,x_N) = \sqrt{\nu_N(x_N)/\hat{\varphi}(N,x_N)}$ in the algorithm to

$$\varphi(N,x_N) = \left(\frac{\nu_N(x_N)}{\hat{\varphi}(N,x_N)}\right)^{\frac{\eta}{\eta+1}}.$$

The convergence rate is strictly bounded from above 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 in view of (12). It is the Markov process with kernel $M(t)$ (assuming that all $M(t)$ are stochastic matrices) and initial distribution ν_0. Indeed, $\frac{\eta}{\eta+1} = 0$ implies $\varphi(N,\cdot) = 1$ on $\mathscr{X}$. In view of (24a), we get $\varphi(t,i) \equiv 1$. This is intuitive and we do not need to run the algorithm to solve the problem when $\eta = 0$.

Example 3. *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*

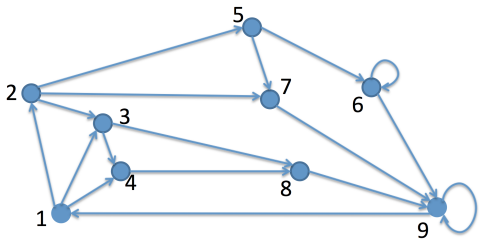


Fig. 1: transport graph

Problem 1 is given by

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.59 & 0.21 & 0.21 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.38 & 0 & 0.21 & 0.41 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.38 & 0 & 0 & 0.62 \end{bmatrix},$$

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 [17]. 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,

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.69 & 0.20 & 0.10 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.10 & 0.10 & 0.49 & 0 & 0.10 & 0.20 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.39 & 0.10 & 0.20 & 0.31 \\ 0 & 0 & 0 & 0 & 0 & 0.29 & 0 & 0 & 0.71 \end{bmatrix}.$$

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, for the evolution of distributions obtained by minimizing the cost in (27) for $\eta = 10$,

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.77 & 0.15 & 0.08 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.08 & 0.08 & 0.61 & 0 & 0.08 & 0.15 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.54 & 0.08 & 0.15 & 0.23 \\ 0 & 0 & 0 & 0 & 0 & 0.46 & 0 & 0 & 0.54 \end{bmatrix},$$

and for $\eta = 100$,

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.78 & 0.15 & 0.07 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.07 & 0.07 & 0.64 & 0 & 0.07 & 0.15 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.57 & 0.07 & 0.14 & 0.22 \\ 0 & 0 & 0 & 0 & 0 & 0.50 & 0 & 0 & 0.50 \end{bmatrix}.$$

2325-5870 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCNS.2019.2935623, IEEE Transactions on Control of Network Systems

As can be seen, the terminal distribution gets closer to ν_N as η increases. However, this comes at a cost of slower convergence. In fact, the number of iterations needed to achieve relative accuracy of 10^{-4} are $12,67,412$, corresponding to η being $1,10,100$, respectively.

The observed phenomenon of "mass-spreading" before final "reassembly" to a distribution close to specification, provides a degree of robustness against link failures, as it increases the alternative ways mass travels to destination. Evidently, any change in topology can evidently be used to update the data (matrices $M(t)$) so as to recompute the transportation scheduling for the remaing of the time window. While it is intuitively clear that the relaxation of the problem leads to a transportation plan the explores a wider range of paths, numerical experiments may be necessary to quantify advantages of such plans on a case by case basis, as the benefits and caveats may depend on the precise topology of the network.

V. Final comments

Since the work of Mikami, Thieullen, Leonard, Cuturi [44], [45], [46], [40], [41], [23], a large number of papers have appeared where Schrödinger bridge problems are viewed as regularization of the Optimal Mass Transport (OMT) problem, see e.g., [7], [42], [2], [19]. Computation of solutions to OMT problems in high dimensional cases is computationally challenging [6], and due to that, the regularization aforded by the Schrödinger bridge problem (aka entropic regularization) present an effective computational scheme to obtain approximate OMT solutions. Besides, Schrödinger bridges are of independent interest as they are motivated by the following two problems: The first is based on Schrödinger's original "hot gas experiment" and seeks *large deviations of the empirical distribution on paths* [27]. The second is a *maximum entropy principle in statistical inference*, namely the problem to choose the *a posteriori* distribution corresponding to 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 [34], [35], [11], [12], [26], [20], [21], [22]. It is this last concept which largely inspired the original approach taken in this paper.

In the setting of the present work, the prior mass distribution on paths may simply encode topological information of the network or, encode topological information together with the length of each link. The mass distribution does not necessarily need to be a probability distribution. Then, given two marginal distributions at the beginning and end of a time interval, we have considered a relaxed version of the Schrödinger problem where the final distribution does not need to be precisely matched, and the solution only needs to correspond to a terminal distribution that is close to the desired one. This has been formalized by adding to the criterion the Kullback-Leibler distance between the final distribution and the desired one. We have shown that the solution can be obtained solving a modified Schrödinger system, in that the terminal matching condition differs from the one in standard Schrödinger problem. Finally, we provided an iterative algorithm, contracting the Hilbert metric with contraction ratio less than $1/2$, to compute the solution numerically.

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] 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.

[4] G. Baggio, A. Ferrante, R. Sepulchre, Finslerian metrics in the cone of spectral densities, ArXiv e-prints, arXiv:1708.02818.

[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] G Birkhoff, Uniformly semi-primitive multiplicative processes, *Transactions of the American Mathematical Society*, **104** (1962), pp. 37-51.

[10] S. Bonnabel, A. Astolfi, and R. Sepulchre, Contraction and observer design on cones, in *Proc. Decision and Control and European Control Conference*, (CDC-ECC), IEEE, 2011, pp. 12-15.

[11] 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.

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

[13] P. Bushell, On the projective contraction ratio for positive linear mappings, *Journal of the London Mathematical Society*, **2**, pp. 256–258, 1973.

[14] P. Bushell, Hilbert's metric and positive contraction mappings in a Banach space, *Archive for Rational Mechanics and Analysis*, **52**, 330-338, 1973.

[15] 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.

[16] 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.

[17] Y. Chen, T.T. Georgiou, M. Pavon and A. Tannenbaum, Robust transport over networks, *IEEE Trans. Aut. Control*, **62**, n.9, 4675-4682, 2017.

[18] Y. Chen, T.T. Georgiou, M. Pavon and A. Tannenbaum, Efficient-robust routing for single commodity network flows, *IEEE Trans. Aut. Control*, **63**, n.7,2287-2294, 2018.

[19] 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.

2325-5870 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCNS.2019.2935623, IEEE Transactions on Control of Network Systems

[20] I. Csiszár, I-divergence geometry of probability distributions and mimimization problems, *Annals of Probability*, **3**, pp. 146-158, 1975.

[21] I. Csiszár, Sanov property, generalized I-projections, and a conditional limit theorem, *Annals of Probability*, **12**, pp. 768-793, 1984.

[22] 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.

[23] M.Cuturi, "Sinkhorn distances: lightspeed computation of pptimal transport," *Advances in Neural Information Processing Systems*, 2292-2300, 2013.

[24] 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.

[25] 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.

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

[27] 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.

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

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

[30] J. Franklin and J. Lorenz, On the scaling of multidimensional matrices, *Linear Algebra and its applications*, **114**, 717-735, 1989.

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

[32] 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.

[33] D. Hilbert, Über die gerade linie als kürzeste verbindung zweier punkte, *Mathematische Annalen*, **46**, 91-96, 1895.

[34] 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.

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

[36] 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

[37] E. Kohlberg and J. W. Pratt, The contraction mapping approach to the Perron-Frobenius theory: Why Hilbert?s metric?, *Mathematics of Operations Research*, **7**, no. 2, 198-210, 1982.

[38] B. Lemmens and R. Nussbaum, *Nonlinear Perron-Frobenius Theory*, no. 189, Cambridge University Press, 2012.

[39] B. Lemmens and R. Nussbaum,, Birkhoff's version of Hilbert's metric and its applications in analysis, ArXiv e-prints, arXiv:1304.7921v1, (2013), *Handbook of Hilbert geometry*, Chapter 10, G. Besson, A. Papadopoulos and M. Troyanov Eds., European Mathematical Society Publishing House, Zürich, 275-306,2014.

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

[41] 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.

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

[43] 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.

[44] 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.

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

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

[47] N. K. Olver, *Robust Network Design*. Ph.D. Dissertation, Department of Mathematics and Statistics, McGill University, 2010.

[48] D. Reeb, M. J. Kastoryano, and M. M. Wolf, Hilbert's projective metric in quantum information theory, *J. Math. Phys.*, **52** (2011), p. 082201.

[49] 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.

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

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

[52] R. Sepulchre, A. Sarlette, and P. Rouchon, Consensus in non-commutative spaces, in *Proc. 49th IEEE-CDC*, pp. 6596-6601, 2010.

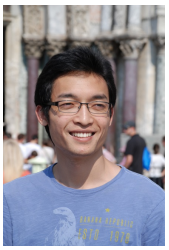
[53] R. Sinkhorn, A relationship between arbitrary positive matrices and doubly stochastic matrices, *Ann. Math. Statist.*, **35** (1964), 876-879.

[54] A. Thompson, On certain contraction mappings in a partially ordered vector space, *Proc. of the American Mathematical Society*, **14**, no. 3, 438-443, 1963.

[55] J. Tsitsiklis, D. Bertsekas, and M. Athans, *Distributed asynchronous deterministic and stochastic gradient optimization algorithms*, *IEEE Transactions on Automatic Control,* **31**, pp. 803–812, 1986.

2325-5870 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCNS.2019.2935623, IEEE Transactions on Control of Network Systems



Yongxin Chen received his BSc from Shanghai Jiao Tong University in 2011 and Ph.D. from University of Minnesota in 2016, both in Mechanical Engineering. He is currently an Assistant Professor in the School of Aerospace Engineering at Georgia Institute of Technology. He has served on the faculty at Iowa State University (2017-2018). He received the George S. Axelby Best Paper Award in 2017 for his joint work with Tryphon Georgiou and Michele Pavon. His current research focuses on the intersection of control theory, machine learning, robotics and optimization.

Tryphon T. Georgiou was educated at the National Technical University of Athens, Greece, and the University of Florida, Gainesville (PhD 1983). He is currently a Chancellor's Professor in the Department of Mechanical and Aerospace Engineering at the University of California, Irvine. He has served on the faculty at Florida Atlantic University (1983-1986), Iowa State University (1986-1989) and the University of Minnesota (1989-2016). Dr. Georgiou is a co-recipient of the George S. Axelby award of the IEEE Control Systems Society for the years 1992, 1999, 2003 and 2017, a Fellow of the Institute of Electrical and Electronic Engineers (IEEE), a Fellow of the International Federation of Automatic Control (IFAC), and a Foreign Member of the Royal Swedish Academy of Engineering Sciences (IVA).

Michele Pavon Michele Pavon was born in Venice, Italy, on October 12, 1950. He received the Laurea degree from the University of Padova, Padova, Italy, in 1974, and the Ph.D. degree from the University of Kentucky, Lexington, in 1979, both in mathematics. After service in the Italian Army, he was on the research staff of LADSEB-CNR, Padova, for six years. Since July 1986, he has been a Professor at the School of Engineering, the University of Padova. Co-recipient (with Y. Chen and T. Georgiou) of the IEEE Control Systems Society 2017 G. S. Axelby Outstanding Paper Award. His present research interests include maximum entropy and optimal transport problems.

Allen Tannenbaum is presently Distinguished Professor of Computer Science and Applied Mathematics/Statistics at SUNY Stony Brook. He works in systems and control, signal processing, computer vision, and systems biology.

2325-5870 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.