Noncommutative Sinkhorn theorem and generalizations

Yongxin Chen, Tryphon Georgiou and Michele Pavon

Our topic traces back to DEMING AND STEPHAN [1] who sought to reconcile an empirical joint distribution of two random variables with *a priori* known marginals for each and proposed an iterative scheme for determining a solution. Thus, in the case of finite probability spaces, the problem data consist of two probability vectors p_0 and p_1 together with the empirical joint probability matrix $P = [p(i,j)]$, i.e., a matrix P such that $p(i,j) \geq 0$ and $\sum_{i,j} p(i,j) = 1$. Then the task is to suitably modify P into $\hat{P} = [p(i,j)]$ which is consistent with the marginals, i.e.,

$$\sum_i \hat{p}(i,j) = p_0(j) \text{ and } \sum_j \hat{p}(i,j) = p_1(i),$$

over the corresponding indexing sets. It took almost twenty years before a proof of convergence for the Deming-Stephan algorithm was provided in [2] and before the problem was fully understood and explained [2], [4], [3].

Closely related work that was motivated by estimation of transition probabilities of Markov chains led to a very similar problem. Here $\Pi = [\pi(i,j)]$ is the transition probability from state i to j of a Markov chain, and the problem which was originally posed by L. Welch (as cited in [5], [6]) is to normalize its entries by alternating between scaling rows and columns, so that they sum up to one, and thereby modify Π so as to become doubly stochastic. Convergence of this algorithm, which is similar to the algorithm of Deming and Stephan, and the fact that under suitable conditions it leads to a doubly stochastic matrix $\hat{\Pi}$ is now known as Sinkhorn's theorem [**?**]. The more general case of the Deming and Stephan's setting corresponds to asking for a transition probability matrix $\hat{\Pi}$ for the Markov chain that satisfies[1]

$$\hat{\Pi}^\dagger p_0 = p_1$$

for the given probability vectors–Sinkhorn's theorem corresponds to the special case where p_0 and p_1 have all their entries equal. For the more general setting the basic theorem can be phrased as follows.

Theorem 1: Let

$$\Pi = [\pi(i,j)]_{i,j=1}^N$$

be a $N \times N$ stochastic matrix i.e., $\pi(i,j) \geq 0$ and $\sum_j p(i,j) = 1$, and in addition assume that all entries are positive ($\pi(i,j) > 0$), and let p_0, p_1 be given probability vectors. Then, there exist a unique pair of diagonal matrices D_1, D_0 such that

$$\hat{\Pi} := D_1 \Pi D_0^{-1}$$

is a stochastic matrix and satisfies

$$\hat{\Pi}^\dagger p_0 = p_1.$$

The solution $\hat{\Pi}$ is seen as a correction of Π so as to agree with the given marginal vectors. The scaling matrices can be obtained by a convergent iteration of scaling columns and rows accordingly. Moreover, it can be shown that this process provides a new law for the Markov chain for which the joint probability law (given by $p(i,j) = \hat{\pi}(i,j)p_0(i)$) between states at two points in time for the Markov chain, is closest in the relative entropy sense to the "prior" law (given by $\pi(i,j)p_0(i)$) and subject to being consistent with the given marginals. For background and further references see [7], [8].

The purpose of the presentation will be to highlight key points of the above theory and the context to discuss a noncommutative generalization. Briefly, this pertains to Quantum Channels [14, Chapter 2] that represent the noncommutative counterparts of Markov transitions. Finite-dimensional noncommutative probability spaces are composed of nonnegative matrices having trace one, herein denoted by

$$\mathfrak{D} := \{\rho \in \mathbb{C}^{n\times n} \mid \rho = \rho^\dagger \geq 0 \text{ and } \operatorname{trace}\rho = 1\}.$$

These are referred to as density matrices and encapsulate the statistics of quantum systems–they are the analog of probability vectors. Linear maps in this category are trace preserving and completely positive (CPTP). Assuming for notation convenience that both domain and range have the same size, CPTP admit the representation

$$\mathcal{E}^\dagger \;:\; \mathfrak{D} \to \mathfrak{D} \;:\; \rho \longrightarrow \sigma = \sum_{i=1} E_i \rho E_i^\dagger, \tag{1a}$$

with $E_i \in \mathbb{C}^{n\times n}$ such that

$$\sum_{i=1} E_i^\dagger E_i = I. \tag{1b}$$

Throughout, I denotes the identity matrix. Condition (1b) ensures that the map preserves the trace, i.e., that $\operatorname{trace}(\sigma) = \operatorname{trace}(\rho) = 1$. *Completely-positive* refers to the property that, besides the fact that $\mathcal{E}^\dagger(\rho) \geq 0$ for all $\rho \geq 0$ (i.e., being a *positive* map), if $\mathbb{I}_k$ denotes the identity map on $\mathbb{C}^{k\times k}$ and $\otimes$

Y. Chen and T.T. Georgiou are with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, Minnesota MN 55455, USA, {chen2468,tryphon}@umn.edu, and M. Pavon is with the Dipartimento di Matematica, Università di Padova, via Trieste 63, 35121 Padova, Italy, pavon@math.unipd.it

The research was supported in part by the NSF under Grant 1027696 and the AFOSR under Grants FA9550-12-1-0319 and FA9550-15-1-0045, and by the University of Padova Research Project CPDA 140897.

[1] Dagger † denotes the adjoint.

the tensor product, then $\mathbb{I}_k \otimes \mathcal{E}^\dagger$ is a positive map for all k. The above so-called Kraus-representation (1a-1b) precisely characterizes the CPTP property (see, e.g., [14, Chapter 2]).

The condition in Theorem 1 that requires strict positivity of transition probabilities ($\pi(i,j) > 0$ for all indices), implies that Π maps probability vectors into the open interior of the probability simplex. Likewise, the following property that we refer to as *positivity improving*,

$$\mathcal{E}^\dagger(\rho) > 0 \text{ for any } \rho \in \mathfrak{D}, \tag{2}$$

amounts to having the image of $\mathcal{E}^\dagger$ contained in the interior of the space of density matrices. Positivity improving can be rephrased as having no pair of vectors $w,\, v \in \mathbb{C}^n$ such that $w^\dagger E_i v = 0$ for all $i \in \{1, \ldots, n_{\mathcal{E}}\}$. Interestingly, while entry-wise positivity of a matrix is easy to ascertain, determining whether a given CPTP map is positivity improving turns out to be NP-hard [9], [11].

While there are many similarities between quantum channels and Markov evolution, there are also fundamental differences. A departure from classical probability arises in that, in general, there is no notion of joint probability between the two ends of a quantum channel. This is due to the fact that the order in which measurements take place matters. For this reason, generalization of the problems of Deming and Stephan, and Sinkhorn for Quantum channels can more naturally be thought in the context of CPTP maps.

Sinkhorn's theorem [5], [6] extends almost vebatim to the noncommutative setting. In this context, the problem amounts to scaling the two ends of the Quantum channel so as modify the channel to be consistent with uniform marginals, i.e., with

$$\rho_0 = \frac{1}{n} I \text{ as well as } \rho_1 = \frac{1}{n} I.$$

We refer to CPTP maps with this property as doubly stochastic. The precise result is:

Theorem 2: Given a positivity improving CPTP map $\mathcal{E}^\dagger_{0:T}$, there exists a pair of square matrices χ_0, χ_1 unique up to scaling by a constant such that

$$\hat{\mathcal{E}}^\dagger(\cdot) := \chi_1 \left(\mathcal{E}^\dagger_{0:T}(\chi_0^{-1}(\cdot)\chi_0^{-\dagger}) \right) \chi_1^\dagger \tag{3}$$

is a *doubly stochastic*, i.e., $\hat{\mathcal{E}}(I) = I$ as well as $\hat{\mathcal{E}}^\dagger(I) = I$.

This result was obtained by Gurvits [10, Theorem 4.7] and independently in [7] using a completely different arguments. In [10] existence was established by constructing certain natural potentials (*locally scalable functionals*) whereas in [7] the proof was based on nonlinear Frobenius-Perron theory and the Hilbert metric.

The full generalization of Theorem 1 to the noncommutative case remains open. This will be the topic of the presentation and it is more precisely stated below as a conjecture (see also [7, Conjecture 1]):

Conjecture 1: Given a positivity improving CPTP map $\mathcal{E}^\dagger_{0:T}$ and two density matrices ρ_0, ρ_1, there exists a pair of square matrices χ_0, χ_1 unique up to scaling by a constant such that

$$\hat{\mathcal{E}}^\dagger(\cdot) := \chi_1 \left(\mathcal{E}^\dagger_{0:T}(\chi_0^{-1}(\cdot)\chi_0^{-\dagger}) \right) \chi_1^\dagger \tag{4}$$

maps ρ_0 into ρ_1 and it is trace preserving, that is, $\hat{\mathcal{E}}^\dagger(\rho_0) = \rho_1$ as well as $\hat{\mathcal{E}}(I) = I$.

In either, Theorem 2 and Conjecture 1, it is seen that $\chi_0\chi_0^\dagger$ and $\chi_1\chi_1^\dagger$ play the role of the scaling matrices D_0 and D_1, respectively. In fact, specialization of the two statements to the commutative case (where, e.g., ρ_0, ρ_1, χ_0, χ_1 are diagonal and the CPTP map reduces to a stochastic matrix) can be seen to correspond precisely to their classical counterpart in Theorem 1. The presentation will highlight the context and overview the theory surrounding this problem.

References

[1] 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. Stat. 11, 427-444 (1940).

[2] S. Fienberg, An iterative procedure for estimation in contingency tables, Ann. Math. Stat. 41, 907-917 (1970).

[3] I.C. Terrance, and S. Kullback, "Contingency tables with given marginals." Biometrika 55.1 (1968): 179-188.

[4] S. Kullback, "Probability densities with given marginals." The Annals of Mathematical Statistics (1968): 1236-1243.

[5] R. Sinkhorn, *A relationship between arbitrary positive matrices and doubly stochastic matrices*, The Annals of Mathematical Statistics, 35 (1964), pp. 876–879.

[6] ———, *Diagonal equivalence to matrices with prescribed row and column sums*, Proceedings of the American Mathematical Society, 45 (1974), pp. 195–198.

[7] Georgiou, T.T., and M. Pavon "Positive contraction mappings for classical and quantum Schrödinger systems." Journal of Mathematical Physics 56.3 (2015): 033301.

[8] Lemmens, Bas, and Roger Nussbaum Nonlinear Perron-Frobenius Theory. No. 189. Cambridge University Press, 2012.

[9] S. Gaubert and Z. Qu, *Checking the strict positivity of Kraus maps is NP-hard*, arXiv preprint arXiv:1402.1429, (2014).

[10] L. Gurvits, *Classical complexity and quantum entanglement*, Journal of Computer and System Sciences, 69 (2004), pp. 448–484.

[11] C. Hillar and L.-H. Lim, *Most tensor problems are NP-hard*, arXiv preprint arXiv:0911.1393, (2009).

[12] C. T. Ireland and S. Kullback, *Contingency tables with given marginals*, Biometrika, 55 (1968), pp. 179–188.

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

[14] D. Petz, *Quantum information theory and quantum statistics*, Springer, 2008.