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/TIT.2013.2286087, IEEE Transactions on Information Theory

Minimum Relative Entropy for Quantum Estimation: Feasibility and General Solution

Mattia Zorzi, Francesco Ticozzi and Augusto Ferrante

***Abstract*—We propose a general framework for solving quantum state estimation problems using the minimum relative entropy criterion. A convex optimization approach allows us to decide the feasibility of the problem given the data, find the maximal common kernel of all admissible states and, whenever necessary, to relax the constraints in order to allow for a physically admissible solution. Building on these results, the variational analysis can be completed ensuring existence and uniqueness of the optimum. The latter can then be computed by standard, efficient standard algorithms for convex optimization, without resorting to approximate methods or restrictive assumptions on its rank.**

***Index Terms*—Quantum estimation, *Kullback-Leibler* divergence, Convex optimization**

I. Introduction

Quantum devices implementing information processing tasks promise potential advantages with respect to their classical counterparts in a remarkably wide spectrum of applications, ranging from secure communications to simulators of large scale physical systems [1], [2], [3].

In order to exploit quantum features to the advantage of a desired task, tremendous challenges are posed to experimentalists and engineers, and many of these have stimulated substantial theoretically-oriented research. Which particular problem is critical depends on the physical system under consideration: from optical integrated circuits to solid-state devices, the tasks in the device engineering, protection from noise and control are manifold [3], [2], [4], [5], [6]. However, *quantum estimation* [7] problems are ubiquitous in applications, be it in testing the output of a quantum algorithm, in reconstructing the behavior of a quantum channel or in retrieving information at the receiver of a communication system [3], [8], [9], [10], [11]. In this paper we focus on state estimation problem for finite-dimensional quantum system, namely the reconstruction of a trace-one, positive semidefinite matrix given from data, and in particular on an estimation method that addresses two critical problems for most real-world situations. The first regards situation in which the system dimension is large and the available measurement are insufficient to completely determine the state. The second arises when only a set of noisy data is available, yielding no physically-acceptable solution.

In order to address the first issue, a typical approach in both the classical and the quantum world is to resort to a MAXENT principle [12], [13], [14], [15], [16], [17], [18], [7], [19], where one opts for a "maximum ignorance" criterion on the choice of parameters that are not uniquely determined by data. The MAXENT estimation can indeed be seen as a particular case of *minimum relative entropy* estimation, [20], where the information-theoretic pseudo-distance of the estimated state with respect to some *a priori* state is minimized subjected to a set of constraints representing the available data [21], [22], [23], [24], [25]. This *a priori* information introduces a new ingredient with respect to typical maximum-likelihood methods for quantum estimation [7], [26], [27], and allows e.g. for taking into account previous estimates of the state to be reconstructed. It is worth remarking that when a full set of observables are not available, any estimation method will not in general be able to reconstruct the actual state. The minimal entropy methods offer a criterion to choose the state that is informationally closest to some a priori guess between those that are compatible with the measurement data.

A quantum minimum relative entropy method for quantum state estimation has been discussed in [23], [28], where *approximate solutions* to minimum relative entropy problems are provided: the estimates are shown to be good approximation of the optimal solution when this is close enough to the a priori state. On the other hand, a way towards the computation of the exact *optimal solution* is indicated in [21]: Georgiou has analyzed the MAXENT problem for estimating positive definite matrices, providing a generic form for the optimal solution, parametric in the Lagrange multipliers. He has also observed that the results can be extended to the more general minimum relative entropy problem.

We shall here extend Georgiou's approach, proving existence, uniqueness and continuity of the solution with respect to the measured data, when a generic prior state is considered. The solution can then be computed by standard numerical methods. However, the approach returns a meaningful answer only when there is a full-rank admissible solution among the states compatible with the data. While this appears to be a reasonable assumption as quantum full-rank states are generic, this is no longer the case whenever the unknown state is pure

Work partially supported by the Italian Ministry for Education and Research (MIUR) under PRIN grant n. 20085FFJ2Z "New Algorithms and Applications of System Identification and Adaptive Control", and by the University of Padua under the QUINTET project of the Dept. of Information Engineering and the QFUTURE strategic Project.

M. Zorzi is with the Department of Electrical Engineering and Computer Science, University of Liège, Institut Montefiore B28, 4000 Liège, Belgium, (mzorzi@ulg.ac.be), A. Ferrante are with the Dipartimento di Ingegneria dell'Informazione, Università di Padova, via Gradenigo 6/B, 35131 Padova, Italy (augusto@dei.unipd.it), and F. Ticozzi is with with the Dipartimento di Ingegneria dell'Informazione, Università di Padova, via Gradenigo 6/B, 35131 Padova, Italy (ticozzi@dei.unipd.it), and the Deptartment of Physics and Astronomy, Dartmouth College, 3127 Wilder, Hanover, NH (USA).

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/TIT.2013.2286087, IEEE Transactions on Information Theory

or near the boundary of the physical state set. In fact, it is easy to picture realistic scenarios where the effect of noisy or biased data might actually force the solutions to be on the boundary, or even outside of the admissible set [29], [26]. In the latter case the constrained optimization problem "as is" is not feasible, and one has to relax the constraints. We here propose a convex-optimization method that solves the feasibility problem and that determines the maximal common kernel of the states compatible with the constraints. This allows for the construction of a reduced optimization problem which can always be successfully tackled within Georgiou's framework.

Regarding the effect of noise and limited accuracy of the available data, our contribution is twofold: we provide both a way to relax the constraints in order to allow for a physically admissible solution when the noisy data are not compatible with physical states, and ensure that the estimated state depends continuously on the data, while providing a way to estimate the error by bounding the modulus of continuity.

The paperis organized as follows: a general setting for posing the feasibility problem and quantum minimum relative entropy with data corresponding to linear constraints is presented in Section II. In Section III, we propose a way to reformulate the *feasibility analysis* as a convex optimization problem. The solution of this ancillary problem, for which we provide a numerical approach in Appendix A, also indicates an optimal way to perturb, or *relax* the constraints in order to allow for admissible solutions to the estimation problem. We also show how the way in which the constraints are relaxed can be tailored to match the error distribution or the level of noise we assume on the measurements. One of the key features of the proposed approach is that, once the feasibility analysis returns a positive answer, it directly leads to the construction of a *reduced problem* for which there exists a positive definite state satisfying the given constraints. In Section IV, we address the corresponding (reduced, if needed) minimum relative entropy problem, showing it admits a unique full-rank solution. The latter can be computed from the closed-form solution of the primal problem, and a standard numerical algorithm to find the corresponding Lagrange multipliers is suggested. Then, the solution to the non-reduced, original problem is immediately obtained. A proof of the continuity of the solution with respect to the data and a way to estimate the induced error are provided in Appendix B. Some concluding remarks and future directions and applications are summarized in Section V.

II. Problem Setting

A. Quantum States and Measurement Data

Consider a quantum n-level system. Its state is described by a density operator, namely by a positive semidefinite unit-trace matrix

$$\rho \in \mathcal{D}_n = \left\{ \rho \in \mathbb{C}^{n\times n} \mid \rho = \rho^\dagger \geq 0,\ \mathrm{tr}(\rho) = 1 \right\}, \quad (1)$$

which plays the role of probability distribution in the classical probability framework. Note that a density matrix depends on $n^2 - 1$ real parameters.

In this work we will be concerned on the problem of reconstructing an unknown ρ from a set of repeated measurement data. This is of course an *estimation* problem in the statistical language, while in the physics community it is usually referred to as *state tomography* [7].

We assume that data are provided in one of the following forms:

1) Outcome frequencies for projective measurements: consider repeated measurements of a (Hermitian) *observable* [30], $O = \sum_k o_k \Pi_k$, where $\{\Pi_k\}$ is the associated spectral family of orthogonal projections. The spectrum $\{o_k\}$ represents the possible outcomes at each measurement, and the frequency of the k-th outcome given a state ρ can be computed as $p_k = \mathrm{tr}(\rho\Pi_k)$. After K measurements of O, we assume we are provided with some experimental estimates of p_k, i.e. the experimental relative frequencies of occurences $\hat{p}_k = \#(O = o_k)/K$, with $\#(O = o_k)$ the number of measurements that returned outcome o_k.

2) Observable averages: consider a set of n_o measured observables, represented by Hermitian matrices O_i, where now we only have access to the mean values of the outcomes, denoted by $\langle O_i \rangle$ (and with possible outcomes $o_{i,k}$), that can be theoretically computed as $\langle O_i \rangle := \mathrm{tr}(\rho O_i)$ and experimentally estimated by $\langle \hat{O}_i \rangle = \sum_k o_{i,k} \hat{p}_k$.

3) Outcome frequencies for general measurements: consider repeated measurements of a *Positive-Operator Valued Measure* (POVM), that is generalized measurements that can be used to describe indirect measurements on a system of interest [3]. A POVM with M outcomes, say $k = 1, \dots M$, is associated with a set of non-negative operators $\{Q_k\}_{k=1}^M$ such that $\sum_k Q_k = I$, playing the role of resolution of the identity for projective measurements. The probability of obtaining the k-th outcome can be computed by $q_k = \mathrm{tr}(\rho Q_k)$, and experimentally estimated by $\hat{q}_k$ after K repeated measurements. This case in fact includes the first one, and the generalization to multiple POVM is straightforward.

In all these scenarios, data are provided as a set of real values representing estimates $\hat{f}_i$ of quantities f_i (that can be either p_i, $\langle O_i \rangle$ or q_i), each associated with the state through a linear relation of the form $f_i = \mathrm{tr}(\rho Z_i)$, where Z_i have the role of Π_i, O_i or Q_i described above. Clearly $\hat{f}_i \to f_i = \mathrm{tr}(Z_i \rho)$ with probability one as $N \to \infty$. Here, N is the number of repeated measurements. This framework is quite general, and can be adapted to include any case if the data are given as linear constraints. Another significant situation that fits in this framework is when reduced states of a multipartite systems are available as data [31], [32]. Finally, by the well-known Choi-Jamiolkowski isomorphism, [33], [34], the same setting, and methods for solution, can be adapted to include estimation of quantum channels, or quantum process tomography [7], [26].

From a theoretical viewpoint, ρ can be in principle reconstructed exactly from at least $n^2 - 1$ averages $f_i = \mathrm{tr}(\rho Z_i)$ $i = 1 \dots n^2 - 1$ when $Z_1 \dots Z_{n^2-1}$ are observables which do not carry redundant information, namely they form a basis for the space of traceless Hermitian matrices.

Remark: In this case, if the measurement statistics f were known exactly, there would be a unique state compatible with

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/TIT.2013.2286087, IEEE Transactions on Information Theory

the corresponding constraints. Hence, the optimization step would become superfluous as the unique state fitting the data would correspond to the actual state. In the case where the measurement statistics $\hat{f}$ are computed from a finite number of data N, the probability that the error is larger than ϵ could be estimated e.g. using the central limit theorem. In fact, $\hat{f}_i$ would be the sample average of N i.i.d. random variables associated to repeated measurements of each observable Z_i on identical preparation of the system. Let σ_i^2 be the variance of such random variables, and assume the Z_i to be orthonormal for simplicity. This allows for the following standard estimate, using Chebyshev's inequality and the law of large numbers:

$$\mathbb{P}(|\hat{f}_i - f_i| \geq \varepsilon) \leq \frac{\sigma_i^2}{N\varepsilon^2}. \tag{2}$$

The estimated solution would have the structure:

$$\hat{\rho} = \frac{1}{2}I + \sum_i f_i Z_i.$$

Therefore, considering the Hilbert-Schmidt norm $\|\cdot\|$, we have $\|\hat{\rho} - \rho\| \leq \sqrt{\sum_i n^2 |\lambda_i|^2 |\hat{f}_i - f_i|^2}$, with λ_i being the eigenvalue of Z_i with maximum absolute value. Define $\varepsilon_{tot} = \sqrt{\sum_i n^2 |\lambda_i|^2 \varepsilon^2}$, in terms of ε used in (2). Hence if we have N measurement available for each observable we can guarantee that

$$\mathbb{P}(\|\hat{\rho} - \rho\| \leq \varepsilon_{tot}) \geq \left(1 - \frac{\sigma_{max}^2}{N\varepsilon^2}\right)^{n^2-1},$$

where σ_{max}^2 is the maximum variance. Given the number of measurements N we can estimate the accuracy of the reconstructed state, which converges to the actual one for N approaching infinity.

In any practical application, however, one has to face the following issues:

1) Accurate estimates $\hat{f}_i$ of f_i are only obtained by averaging over a large quantity of trials; Often only a small set of trials is available, and/or the data are subject to significant errors;
2) The number of observables required for a unique reconstruction of ρ grows quadratically with respect to the dimension of the quantum system, and exponentially in the number of subsystems. Typically only a small subset of these is available;

We here analyze the estimation problem when these two aspects are taken into account. The first one will lead us to consider the *feasibility* problem, that is, if the problem admits a physically admissible solution for the given data. Since errors may affect the $\hat{f}_i$, the reconstructed state may not be positive semidefinite, or a valid state that satisfies the constraints might not even exist. The second issue generically leads to a estimation problem where more than one state satisfy the constraints, and thus an additional criterion has to be introduced to arrive at a unique solution. As we said, a typical strategy in this setting is to introduce an entropic functional, e.g. *relative entropy* with respect to some reference state representing *a priori* information.

B. Statement of the Main Problems

Consider the setting described above, where we want to estimate the state of an n-dimensional quantum systems from the real data $\{\hat{f}_i\}_{i=2}^p$, experimental estimates of the quantities $f_i = \mathrm{tr}(Z_i\rho)$, for the Hermitian matrices $Z_2 \dots Z_p$, with $p \ll n^2 - 1$. In addition to these, we introduce an auxiliary observable $Z_1 = I$ and the corresponding estimate $\hat{f}_1 = 1$. In this way, we include the linear constraint $\mathrm{tr}(\rho) = 1$ in the constraints associated with the "data". We wish now to solve the following problem.

Problem 1: Given $\{Z_i\}$ and $\{\hat{f}_i\}$, $i = 1 \dots p$, find all:

$$\rho \in \mathcal{H}_n, \ \text{ such that } \ \rho \geq 0, \ \ \hat{f}_i = \mathrm{tr}(\rho Z_i), \ i = 1, \dots, p. \tag{3}$$

Here, $\mathcal{H}_n$ denotes the vector space of Hermitian matrices of dimension equal to n. Notice that, if we remove the positivity constraint $\rho \geq 0$, all other constraints are linear and identify a hyperplane in $\mathcal{H}_n$. To our aim it is convenient to first address a simpler problem: let

$$\mathcal{S} := \left\{\rho \in \mathcal{H}_n \mid \rho \geq 0, \ \ \hat{f}_i = \mathrm{tr}(\rho Z_i)\right\}$$

be the set of the density matrices which solve Problem 1.

Problem 2 (Feasibility): Determine if $\mathcal{S}$ is not empty.

When the problem is feasible, in general $\mathcal{S}$ contains more than one solution, and in principle any solution in $\mathcal{S}$ fits the data. We focus on choosing a solution that has minimum distance with respect to an *a priori* state [1]. In the same spirit of MAXENT problem, this corresponds to give maximum priority to fitting the data, and then choosing the admissible solution that is the closest (in the relative-entropy pseudo-distance) to our *a priori* knowledge on the systems. To this aim, consider the (Umegaki's) quantum relative entropy between $\rho \in \mathcal{D}_n$, and $\tau \in \mathcal{D}_n$ [35]:

$$\mathbb{S}(\rho\|\tau) = \mathrm{tr}(\rho \log \rho - \rho \log \tau). \tag{4}$$

If we assume $\tau \in \mathrm{int}(\mathcal{D}_n)$ and the usual convention that $0 \log(0) = 0$, we do not have to worry about unbounded values of $\mathbb{S}(\rho\|\tau)$. We will see that the condition $\tau \in \mathrm{int}(\mathcal{D}_n)$ is not necessary if ρ lies on the boundary of the admissible set, but whether τ is acceptable depends on the data: this will be discussed in Section IV-A.

Problem 3 (Minimum relative entropy estimation): Given the observables $Z_1 \dots Z_p$, the corresponding estimates $\hat{f}_1 \dots \hat{f}_p$ and a state $\tau \in \mathrm{int}(\mathcal{D}_n)$, solve

$$\underset{\rho \geq 0}{\text{minimize}} \ \mathbb{S}(\rho\|\tau) \ \text{ subject to } \ \mathrm{tr}(\rho Z_i) = \hat{f}_i, \ i = 1 \dots p. \tag{5}$$

Here, τ represents the *a priori* information on the considered quantum system. We set $\tau = \frac{1}{n}I$ if no information is available. In this situation, $\mathbb{S}(\rho\|\tau) = \log(n) - \mathbb{S}(\rho)$, where $\mathbb{S}(\rho) = -\mathrm{tr}(\rho \log(\rho))$ is von Neumann's entropy of ρ, and thus the problem is equivalent to a MAXENT problem. Note that, in either case, the solution to the Problem above may be singular.

[1] It is worth noting that, in a Bayesian perspective, this correspond to assuming a *prior* distribution on the set of states which is a Dirac delta centered in τ, i.e. in absence of additional information the previously available estimate is considered the best one.

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/TIT.2013.2286087, IEEE Transactions on Information Theory

III. Feasibility Analysis

A. An auxiliary problem

We start by addressing the *feasibility* problem, i.e. to determine when $\mathcal{S}$ is not empty. In addition, whenever the problem is not feasible, we show how to determine a suitable perturbation of the $\{\hat{f}_i\}$ that makes our problem feasible. We will show that the corresponding $\mathcal{S}$ only contains singular density matrices when the constraints are relaxed.

Note that, the constraints are linear in $\hat{f}_i$ and Z_i, and can be linearly combined: $\alpha\hat{f}_i + \beta\hat{f}_k = \mathrm{tr}[\rho(\alpha Z_i + \beta Z_k)]$ for each $\alpha, \beta \in \mathbb{R} \setminus \{0\}$ and $i, k = 1 \dots p$. Consider the vector space generated by the observed operators, $\mathrm{span}\{Z_1 \dots Z_p\}$. Thus, by applying the *Gram-Schmidt* process, starting with $X_1 = \frac{1}{\sqrt{n}} Z_1 = \frac{1}{\sqrt{n}} I$, we can compute an orthonormal basis for it:

$$X_i := \alpha_i^i Z_i + \sum_{l=1}^{i-1} \alpha_l^i X_l, \quad i = 2 \dots m \leq p. \tag{6}$$

By linearity, by associating these basis elements to the estimates

$$\begin{aligned} \bar{f}_1 &:= \frac{1}{\sqrt{n}} \\ \bar{f}_i &:= \alpha_i^i \hat{f}_i + \sum_{l=1}^{i-1} \alpha_l^i \bar{f}_l, \quad i = 2 \dots m, \end{aligned} \tag{7}$$

we obtain a new yet *equivalent* set of constraints:

$$\bar{f}_i = \mathrm{tr}(\rho X_i), \quad i = 1 \dots m. \tag{8}$$

Note that, $I \in \mathrm{span}\{X_1 \dots X_m\}$.

Let $Y_1 \dots Y_{n^2-m}$ be an orthonormal completion of $X_1 \dots X_m$ to a basis of $\mathcal{H}_n$. Accordingly, all the Hermitian matrices, and in particular density operators, can be expressed as

$$\rho = \sum_{i=1}^{m} \alpha_i X_i + \sum_{i=1}^{n^2-m} \beta_i Y_i, \tag{9}$$

with $\mathrm{tr}(\rho X_i) = \alpha_i$. In particular, all the Hermitian matrices satisfying the linear constraints in (8) depend on $n^2 - m$ parameters $\beta = [\beta_1 \dots \beta_{n^2-m}]$:

$$\rho = \tilde{\rho}_0 + \sum_{i=1}^{n^2-m} \beta_i Y_i \tag{10}$$

where we have defined the (not necessarily positive) pseudo-state associated with the constraints:

$$\tilde{\rho}_0 = \sum_{i=1}^{m} \bar{f}_i X_i. \tag{11}$$

In the light of this observation, the feasibility problem consists in checking if there exists at least one vector $\beta \in \mathbb{R}^{n^2-m}$ such that $\rho \geq 0$. To this aim, we introduce an auxiliary problem. Intuitively, the idea is the following: given any Hermitian matrix $\tilde{\rho}_0$, there always exists a real μ such that $\tilde{\rho}_0 + \mu I$ is positive definite. More precisely, if $\tilde{\rho}_0$ is not positive definite already, it is easy to see that such a μ will need to be positive. On the other hand, if $\tilde{\rho}_0$ is already positive, the perturbed matrix remains positive semi-definite for some small, negative μ. Studying the minimal μ that correspond to a positive semidefinite matrix offers us a way to understand whether our constraints allow for physically admissible solutions.

Let us formalize these idea: we define $c := \begin{bmatrix} 0 & \dots & 0 & 1 \end{bmatrix}^T \in \mathbb{R}^{n^2-m+1}$, and

$$H(\underline{v}) := \tilde{\rho}_0 + \sum_{i=1}^{n^2-m} v_i Y_i + v_{n^2-m+1} X_1$$

with $\underline{v} = \begin{bmatrix} v_1 & \dots & v_{n^2-m+1} \end{bmatrix}^T$ and we consider the following minimum eigenvalue problem.

Problem 4: Given $\tilde{\rho}_0$ as in (11) and $Y_1 \dots, Y_{n^2-m}$ an orthonormal completion of $\mathrm{span}\{X_1 \dots X_m\}$, solve

$$\text{minimize } c^T \underline{v} \quad \text{subject to} \quad \underline{v} \in \mathcal{I} := \{\underline{v} \mid H(\underline{v}) \geq 0\}. \tag{12}$$

Lemma 3.1: Problem 4 always admits solution.

Proof: First of all, notice that Problem 4 is a convex optimization problem, and the objective function $c^T \underline{v}$ is linear and continuous over the set $\mathcal{I}$. Then, the proof is divided in three steps.

Step 1: We show that $c^T \underline{v} = v_{n^2-m+1}$ is *bounded from below* on $\mathcal{I}$: since $X_1, \dots, X_m, Y_1, \dots, Y_{n^2-m}$ forms an orthonormal basis and $I \in \mathrm{span}\{X_1, \dots, X_m\}$, the matrices $\{Y_i\}$ are traceless. Thus,

$$\begin{aligned} \mathrm{tr}[H(\underline{v})] &= \mathrm{tr}[\tilde{\rho}_0 + \sum_{i=1}^{n^2-m} v_i Y_i + v_{n^2-m+1} X_1] \\ &= \mathrm{tr}(\tilde{\rho}_0) + \sqrt{n} v_{n^2-m+1} = 1 + \sqrt{n} v_{n^2-m+1} \end{aligned}$$

and $\mathrm{tr}[H(\underline{v})] \geq 0$ for each $\underline{v} \in \mathcal{I}$. Hence, $c^T \underline{v} = v_{n^2-m+1} \geq -\frac{1}{\sqrt{n}}$ for each $\underline{v} \in \mathcal{I}$.

Step 2: Let us consider $\underline{v}_0 = \begin{bmatrix} 0 & \dots & 0 & \sqrt{n}(-\lambda_{\min}(\tilde{\rho}_0) + 1) \end{bmatrix} \in \mathcal{I}$ where $\lambda_{\min}(\tilde{\rho}_0)$ denotes the minimum eigenvalue of $\tilde{\rho}_0$. Accordingly, $c^T \underline{v}_0 = \sqrt{n}(-\lambda_{\min}(\tilde{\rho}_0) + 1)$ and Problem 4 is equivalent to minimize $c^T \underline{v}$ over the closed sublevel set $\mathcal{I}_0 = \{\underline{v} \mid H(\underline{v}) \geq 0, \quad -\frac{1}{\sqrt{n}} \leq v_{n^2-m+1} \leq \sqrt{n}(-\lambda_{\min}(\tilde{\rho}_0) + 1)\} \subset \mathcal{I}$. We want to show that $\mathcal{I}_0$ is *bounded* and accordingly *compact* (recall that we are working in a finite dimensional space). This can done by proving that a sequence $\left\{\underline{v}^k\right\}_{k \geq 0}$ such that $\|\underline{v}^k\| \to \infty$ cannot belong to $\mathcal{I}_0$. It is therefore sufficient to show that the minimum eigenvalue of the associated Hermitian matrix $H(\underline{v}^k)$ tends to $-\infty$ as $\|\underline{v}^k\| \to \infty$ with v_{m^2-n+1} bounded. Note that the affine map $\underline{v} \mapsto H(\underline{v})$ is injective, since $Y_1 \dots Y_{n^2-m}, X_1$ are linearly independent. Accordingly $\|H(\underline{v}^k)\| \to \infty$ as $\|\underline{v}^k\| \to \infty$. Since $H(\underline{v}^k)$ is an Hermitian matrix, $H(\underline{v}^k)$ has an eigenvalue η_k such that $|\eta_k| \to \infty$ as $\|\underline{v}^k\| \to \infty$. By construction $\mathrm{tr}[H(\underline{v}^k)] = 1 + \sqrt{n} v^k_{n^2-m+1}$ and $v^k_{n^2-m+1}$ is bounded in $\mathcal{I}_0$. Thus $\mathrm{tr}[H(\underline{v}^k)] < \infty$, namely the sum of its eigenvalues is always bounded. Thus, there exists an eigenvalue of $H(\underline{v}^k)$ which approaches $-\infty$ as $k \to \infty$. So, $\mathcal{I}_0$ is bounded.

Step 3: Since $c^T \underline{v}$ is continuous over the compact set $\mathcal{I}_0$, by Weiestrass' theorem we conclude that $c^T \underline{v}$ admits a minimum point over $\mathcal{I}_0$. ∎

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/TIT.2013.2286087, IEEE Transactions on Information Theory

We need to take into account that the vector which minimizes $c^T \underline{v}$ over $\mathcal{I}$ may not be in general unique. However, to our aim we are more interested in the sign of the minimum.

Proposition 3.1: Let $\mu = \min_{\underline{v} \in \mathcal{I}} c^T \underline{v}$. Then, the following facts hold:

1) If $\mu > 0$, then Problem 1 is not feasible
2) If $\mu < 0$, then Problem 1 is feasible and there exists at least one positive definite matrix satisfying constraints in (8)
3) If $\mu = 0$, then Problem 1 is feasible and all the matrices satisfying constraints in (8) are singular.

Proof: Note that $G(v_1, \dots, v_{n^2-m}) := \tilde{\rho}_0 + \sum_{i=1}^{n^2-m} v_i Y_i$ represents the parametric family of Hermitian matrices (not necessary positive semidefinite) satisfying constraints in (8). Define $\varepsilon := -v_{n^2-m+1}$, thus Problem 4 can be rewritten in the following way:

$$\text{maximize } \varepsilon \text{ subject to } G(v_1, \dots, v_{n^2-m}) \geq \frac{\varepsilon}{\sqrt{n}} I. \quad (13)$$

Let $\varepsilon^\circ = -\mu$ be the solution of the above problem. If $\varepsilon^\circ < 0$, the parametric family $G(v_1, \dots, v_{n^2-m})$ does not contains positive semidefinite matrices, accordingly Problem 1 does not admit solution. If $\varepsilon^\circ \geq 0$, the parametric family $G(v_1, \dots, v_{n^2-m})$ contains at least one positive semidefinite matrix and Problem 1 admits solution. Moreover if $\varepsilon^\circ > 0$, the parametric family contains at least one matrix $\rho \geq \frac{\varepsilon^\circ}{\sqrt{n}} I > 0$ which is positive definite. On the contrary, for $\varepsilon^\circ = 0$ there only exist positive semidefinite matrices which are singular. ■

An effective numerical approach for the solution to the problem is described in Appendix A. This also allows to determine the maximal common kernel of the states in $\mathcal{S}$, which is key to the general solution we present in Section IV.

In the light of the previous result, if $\mu < 0$ Problem 1 is feasible and $\mathcal{S}$ contains at least one positive definite solution. As we will see in Section IV, this condition ensures that a minimum relative entropy criterion will lead to an admissible solution in $\mathcal{S}$. The remaining cases need to be studied more carefully. We start by showing how to make Problem 1 feasible when it is not be so for the given constraints. It turns out that a minimally relaxed problem is feasible and $\mathcal{S}$ only contains singular density matrices. Next, we deal with the case in which Problem 1 is feasible and all its solutions are singular, showing how they all share a minimal kernel and how to construct a reduced problem with a full-rank solution for which the minimum relative entropy methods work.

B. Forcing the feasibility condition (case $\mu > 0$)

The parameter μ given by the auxiliary problem described above reveals if the original problem is feasible, but also suggests an "optimal" way to relax unfeasible constraints so that they make Problem 1 feasible. In fact, from the definition of μ, we know that there exist $v_1 \dots v_{n^2-m} \in \mathbb{R}$ such that

$$\tilde{\rho}_\mu := \tilde{\rho}_0 + \sum_{i=1}^{n^2-m} v_i Y_i + \mu X_1 \geq 0,$$

and:

$$\operatorname{tr}(\tilde{\rho}_\mu) = 1 + \sqrt{n}\mu.$$

From this positive operator, in order to obtain a density operator, we only need to normalize the trace by defining:

$$\begin{aligned} \rho &:= \frac{1}{1+\sqrt{n}\mu} \tilde{\rho}_\mu \quad (14) \\ &= \frac{1}{\sqrt{n}} X_1 + \sum_{i=2}^{m} \frac{\bar{f}_i}{1+\sqrt{n}\mu} X_i + \sum_{i=1}^{n^2-m} \frac{v_i}{1+\sqrt{n}\mu} Y_i. \end{aligned}$$

This implies that the original problem can be made feasible by uniformly, "isotropically" *contracting* the data $\{\bar{f}_i\}$ of a factor $1/(1+\sqrt{n}\mu)$ and, in light of the fact that μ is a solution to Problem 4, that this is the *minimum* amount of contraction that makes Problem 1 feasible. Moreover, the corresponding set $\mathcal{S}$ only contains singular solutions.

However, the entries in the data set $\{\bar{f}_i\}$ may differ in their reliability, and one would like to be able to relax the corresponding constraints accordingly. This is complicated by the fact that the original $\{Z_i\}$ may not be orthogonal, and the data we are contracting are in fact the linearly transformed output of the Gram-Schmidt orthonormalization described above.

This weighed relaxation can be realized as follows: consider the initial setting of Section II, where we have p observables $Z_1 \dots Z_p$ (not necessarily orthonormal), with $Z_1 = I$ and $\hat{f}_1 = 1$. Define the *reliability indexes* $0 < d_2 \dots d_p \leq 1$ associated with each observable $Z_2 \dots Z_p$. More precisely, the more $\hat{f}_i$ is reliable, the closer to one d_i is. This information can be extracted, for example, from an error analysis on the measurement procedures, with d_i associated with the normalized reciprocal of the variances.

When we obtain the orthonormal generators $X_1 \dots X_m$, the *Gram-Schmidt* process induces a linear transformation on the original estimates $\hat{f}_1, \dots, \hat{f}_p$:

$$\begin{bmatrix} \bar{f}_1 \\ \vdots \\ \bar{f}_m \end{bmatrix} = T \begin{bmatrix} \hat{f}_1 \\ \vdots \\ \hat{f}_p \end{bmatrix} \quad (15)$$

where $T = \begin{bmatrix} \frac{1}{\sqrt{n}} & 0 \\ T_1 & T_2 \end{bmatrix} \in \mathbb{R}^{m \times p}$.

In order to modify the data $\{\hat{f}_i\}$ according to their reliability indexes, we define the new set of data:

$$\begin{bmatrix} \hat{f}'_1 \\ \vdots \\ \hat{f}'_m \end{bmatrix} = T \begin{bmatrix} \hat{f}_1 \\ k d_2 \hat{f}_2 \\ \vdots \\ k d_p \hat{f}_p \end{bmatrix}. \quad (16)$$

where $k > \max\{d_i^{-1}\}$. In this way $\hat{f}_2, \dots, \hat{f}_p$ are amplified of a factor $k d_i > 1$ according their reliability indexes. This will allow for the maximum contraction to be applied to the most noisy estimates.

In order to compute the *minimum* μ that makes the original problem feasible perturbing the data consistently with their

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/TIT.2013.2286087, IEEE Transactions on Information Theory

reliability indexes, we can solve Problem 4 with respect to the new pseudo-state:

$$\tilde{\rho}'_0 = \sum_{i=1}^{m} \hat{f}'_i X_i. \tag{17}$$

It is easy to see that if the original problem was unfeasible, this modified problem is unfeasible as well for k large enough, since all the $\hat{f}_i$ corresponding to traceless operators are multiplied for a factor $kd_i >> 1$. Let $\mu' > 0$ be the parameter given by the auxiliary problem when $\tilde{\rho}'_0$ is considered. By the results of the previous subsection and (14) above, we consider the perturbed constraints

$$\begin{aligned} \bar{f}_1 &= \frac{1}{\sqrt{n}} \\ \bar{f}_i &= \frac{1}{1+\sqrt{n}\mu'} \hat{f}'_m, \quad i = 2 \dots m. \end{aligned} \tag{18}$$

Thus, the corresponding Problem 1 is feasible and $\mathcal{S}$ only contains singular solutions.

Observation: Let us consider the limit case in which some $d_i = 0$. One immediate way of dealing with it would be to exclude the corresponding data: this implies that in our estimation procedure the number of independent constraints is decreased, and hence the measured information associated to the expectation of Z_i is completely discarded. The component of the estimated state along the corresponding (traceless) component of Z_i is then computed in the optimization step by minimizing the relative entropy with respect to τ.

However, one may consider a different approach, in which the unreliable measurement still has an effect. More precisely, consider the estimate $\hat{f}_i$ associated with the observable Z_i and with the smallest reliability index $d_i > 0$. Accordingly, $k = \alpha d_i^{-1}$ with $\alpha > 1$ and $\mathrm{tr}(\tilde{\rho}'_0 Z_i) = \alpha \hat{f}_i$. Let Z_i^I and $Z_i^\perp$ be the unique decomposition $Z_i = Z_i^I + Z_i^\perp$ such that $\mathrm{tr}(Z_i^\perp) = 0$ and $\mathrm{tr}(Z_i^I Z_i^\perp) = 0$. Moreover, let $\hat{f}_i^I$ and $\hat{f}_i^\perp$ be such that $\mathrm{tr}(\tilde{\rho}'_0 Z_i^I) = \hat{f}_i^I$ and $\mathrm{tr}(\tilde{\rho}'_0 Z_i^\perp) = \hat{f}_i^\perp$. The corresponding density operator $\rho := \frac{1}{1+\sqrt{n}\mu} \tilde{\rho}_\mu$ is such that $\mathrm{tr}(\rho Z_i^\perp) = \alpha \frac{\hat{f}_i^\perp}{1+\sqrt{n}\mu}$. Hence, in the limit $d_i \to 0$ we get $\mu \to \infty$ and thus $\mathrm{tr}(\rho Z_i^\perp) \to 0$.

This is equivalent to the expectation that the maximally mixed state would have along $Z_i^\perp$, and can be interpreted in our estimation framework as follows: a completely unreliable measurement has been performed, and its effect is to "substitute" the information associated to the sample average of $Z_i^\perp$ with the one corresponding to maximum ignorance (MAXENT), that is, the same we would obtain if the state were the maximally mixed state.

C. Case $\mu = 0$

In the limit case $\mu = 0$, not only all solutions are singular, but they share a key property.

Proposition 3.2: Assume that, with the definition above, $\mu = 0$. Then there exists a kernel $\mathcal{K}$ which is common for all $\rho \in \mathcal{S}$.

Proof: Let us assume $\mu = 0$, accordingly Problem 1 does only admit singular solutions, with $\dim \ker(\rho) > 0$ $\forall\ \rho \in \mathcal{S}$. Pick a solution $\rho^\circ \in \mathcal{S}$ with kernel of minimal dimension. Suppose by contradiction that there exists $\bar{\rho} \in \mathcal{S}$ such that $\ker(\rho^\circ) \not\subseteq \ker(\bar{\rho})$. Taking into account $p \in (0,1)$, we define $\rho := p\rho^\circ + (1-p)\bar{\rho} \in \mathcal{S}$. Accordingly $\dim \ker(\rho) < \dim \ker(\rho^\circ)$ which is a contradiction, since ρ° has kernel with minimal dimension on $\mathcal{S}$. We conclude that $\mathcal{K} = \ker(\rho^\circ) \subseteq \ker(\rho)\ \forall\ \rho \in \mathcal{S}$. ■

This directly implies the following block-form for all the solutions to Problem 1.

Corollary 3.1: Let $\rho^\circ \in \mathcal{S}$ be a solution with minimal kernel $\mathcal{K}$ and consider its block-diagonal form

$$\rho^\circ = U \begin{bmatrix} \rho_1^\circ & 0 \\ 0 & 0 \end{bmatrix} U^\dagger,$$

where U is a unitary change of basis consistent with the Hilbert space decomposition $\mathcal{H} = \mathcal{K}^\perp \oplus \mathcal{K}$ so that $\rho_1^\circ > 0$. Then, the set of all the solutions of Problem 1 is

$$\mathcal{S} = \left\{ \rho = U \begin{bmatrix} \rho_1 & 0 \\ 0 & 0 \end{bmatrix} U^\dagger \mid \rho_1 \geq 0,\ \bar{f}_i = \mathrm{tr}(\rho X_i) \right\}. \tag{19}$$

As consequence of Corollary 3.1, we can focus on a reduced version of Problem 1, by considering optimization only on the support of ρ°, for which the minimum relative entropy is applicable since ρ_1° is positive definite, see Section IV.

IV. State Estimation with Minimum Relative Entropy Criterion

A. Reduced Problem

In the previous part of the paper we showed how to check the feasibility of Problem 1 given the constraints associated with the data and, if needed, how to relax the constraints in such a way that the corresponding Problem 1 is feasible. In general, however, the set of solutions $\mathcal{S}$ is not constituted by only one element. In this section, we show how to choose, and then compute, a solution in $\mathcal{S}$ according the *minimum quantum relative entropy* criterion.

Given the results of the previous sections, we can assume that either Problem 1 admits at least one (strictly) positive definite solution, or we can resort to a reduced problem for which a full rank solution exists. In fact, if $\mathcal{S}$ only contains singular matrices (case $\mu = 0$, or after relaxation of the constraints), by Corollary 3.1 we have that the set of solution is

$$\mathcal{S} = \left\{ \rho = U \begin{bmatrix} \rho_1 & 0 \\ 0 & 0 \end{bmatrix} U^\dagger \mid \rho_1 \geq 0,\ \bar{f}_i = \mathrm{tr}(\rho X_i) \right\} \tag{20}$$

for some unitary change of basis U consistent with the Hilbert space partition $\mathcal{H} = \mathcal{K}^\perp \oplus \mathcal{K}$. Accordingly for each $\rho \in \mathcal{S}$, constraints in (8) can be rewritten in the following way

$$\bar{f}_i = \mathrm{tr}(\rho X_i) = \mathrm{tr}\left(U \begin{bmatrix} \rho_1 & 0 \\ 0 & 0 \end{bmatrix} U^\dagger X_i \right) = \mathrm{tr}(\rho_1 \bar{X}_i) \tag{21}$$

where $\bar{X}_i := \begin{bmatrix} I & 0 \end{bmatrix} U^\dagger X_1 U \begin{bmatrix} I \\ 0 \end{bmatrix} \in \mathcal{H}_{n_1}$ with $n_1 < n$. Accordingly Problem 1 is equivalent to the corresponding reduced problem with $\bar{X}_1 \dots \bar{X}_m$ and $\bar{f}_1 \dots \bar{f}_m$. The corresponding set of solutions is

$$\mathcal{S}_1 = \left\{ \rho_1 \in \mathcal{H}_{n_1} \mid \rho_1 \geq 0,\ \bar{f}_i = \mathrm{tr}(\rho_1 \bar{X}_i) \right\} \tag{22}$$

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/TIT.2013.2286087, IEEE Transactions on Information Theory

which contains the positive definite solution ρ_1°. Once chosen a solution $\hat{\rho}_1 \in \mathcal{S}_1$, the corresponding original solution is

$$\hat{\rho} = U \begin{bmatrix} \hat{\rho}_1 & 0 \\ 0 & 0 \end{bmatrix} U^\dagger. \tag{23}$$

Finally, given the *a priori* estimate $\tau \in \mathrm{int}(\mathcal{D}_n)$ it is easy to see that $\mathbb{S}(\rho\|\tau) = \mathbb{S}(\rho_1\|\tau_1)$ where

$$\tau_1 := \begin{bmatrix} I & 0 \end{bmatrix} U^\dagger \tau U \begin{bmatrix} I \\ 0 \end{bmatrix} \in \mathrm{int}(\mathcal{D}_{n_1}).$$

Notice that in order to have bounded values of the entropy it is only necessary that $\mathrm{range}(\rho) \subseteq \mathrm{range}(\tau)$. However, whether a certain τ is acceptable can be determined only after having analyzed the data and found the minimal kernel $\mathcal{K}$. In order to avoid this complication, it is convenient to choose a $\tau \in \mathrm{int}(\mathcal{D}_n)$ from the beginning. Full rank τ can be chosen arbitrarily close to any prior state with nontrivial kernel. In the effort of keeping a simple notation, we will not distinguish between the reduced and the full problem in the following discussion, therefore using ρ for either the full or the reduced state, τ for either the full or the reduced *a priori* state, $\{X_i\}$ for either the full or reduced observable, and n for the dimension of the full Hilbert space or the reduced one as needed. We can consider the following simpler problem, restricted to strictly positive matrices:

Problem 5: Given $\tau \in \mathrm{int}(\mathcal{D}_n)$, the observables $X_1 \dots X_m$ and the corresponding estimates $\bar{f}_1 \dots \bar{f}_m$, solve

$$\underset{\rho>0}{\text{minimize}}\ \mathbb{S}(\rho\|\tau) \text{ subject to } \mathrm{tr}(\rho \bar{X}_i) = \bar{f}_i,\ i = 1 \dots m. \tag{24}$$

B. Lagrangian and Form of the Full-Rank Solution

Now we are ready to derive a solution method for problem the entropic criterion. Consider the linear operator associated with the above constraints:

$$\begin{aligned} L \quad &: \quad \mathcal{H}_n \to \mathbb{R}^m \\ &\qquad \rho \mapsto \begin{bmatrix} \mathrm{tr}(\rho X_1) \\ \vdots \\ \mathrm{tr}(\rho X_m) \end{bmatrix}. \end{aligned} \tag{25}$$

Given $\lambda = \begin{bmatrix} \lambda_1 & \dots & \lambda_m \end{bmatrix}^T \in \mathbb{R}^m$ and $\rho \in \mathcal{H}_n$,

$$\langle L(\rho), \lambda \rangle = \sum_{i=1}^m \lambda_i \,\mathrm{tr}(\rho X_i) = \mathrm{tr}(\rho \sum_{i=1}^m \lambda_i X_i) = \langle \rho, L^*(\lambda) \rangle \tag{26}$$

where

$$\begin{aligned} L^* \quad &: \quad \mathbb{R}^m \to \mathcal{H}_n \\ &\qquad \lambda \mapsto \sum_{i=1}^m \lambda_i X_i \end{aligned} \tag{27}$$

is the adjoint operator of L. Define $\bar{f} = \begin{bmatrix} \bar{f}_1 & \dots & \bar{f}_m \end{bmatrix}^T$. Since Problem 5 is a constrained convex optimization problem, we consider its Lagrangian

$$\begin{aligned} \mathcal{L}(\rho, \lambda) &= \mathrm{tr}(\rho \log \rho - \rho \log \tau) - \left\langle \lambda, \bar{f} - L(\rho) \right\rangle \\ &= \mathrm{tr}(\rho \log \rho - \rho \log \tau) + \langle L^*(\lambda), \rho \rangle - \left\langle \lambda, \bar{f} \right\rangle \\ &= \mathrm{tr}[\rho(\log \rho - \log \tau + L^*(\lambda))] - \left\langle \lambda, \bar{f} \right\rangle \end{aligned} \tag{28}$$

where $\lambda \in \mathbb{R}^m$ is the Lagrange multiplier. Note that $\mathcal{L}(\cdot, \lambda)$ is strictly convex over $\mathcal{H}_{n,+}$ where $\mathcal{H}_{n,+}$ denotes the cone of the positive definite matrices. Thus, its minimum point is given by annihilating its first variation

$$\delta\mathcal{L}(\rho, \lambda; \delta\rho) = \mathrm{tr}[(\log \rho + I - \log \tau + L^*(\lambda))\delta\rho] \tag{29}$$

for each direction $\delta\rho \in \mathcal{H}_n$. Accordingly, the unique minimum point for $\mathcal{L}(\cdot, \lambda)$ is

$$\rho(\lambda) = \mathrm{e}^{\log \tau - I - L^*(\lambda)} \tag{30}$$

and

$$\mathcal{L}(\rho(\lambda), \lambda) \leq \mathcal{L}(\bar{\rho}, \lambda),\ \forall\ \bar{\rho} \in \mathcal{H}_{n,+}. \tag{31}$$

If there exists λ° such that $\rho(\lambda^\circ) \in \mathcal{S}$, i.e. $\bar{f} = L(\rho(\lambda^\circ))$, then (31) implies

$$\mathbb{S}(\rho(\lambda^\circ)\|\tau) \leq \mathbb{S}(\bar{\rho}\|\tau),\ \forall\ \bar{\rho} \in \mathcal{S}. \tag{32}$$

Thus, if we are able to find $\lambda^\circ \in \mathbb{R}^m$ such that

$$\bar{f} - L(\rho(\lambda^\circ)) = 0, \tag{33}$$

then $\rho(\lambda^\circ)$ is the unique solution to Problem 5. This issue can be addressed by considering the dual problem. In fact, λ° is a solution of (33) if and only if it maximizes the following dual functional over $\mathbb{R}^m$

$$\begin{aligned} \inf_{\rho \in \mathcal{H}_{n,+}} \mathcal{L}(\rho, \lambda) &= \mathcal{L}(\rho(\lambda), \lambda) \\ &= -\mathrm{tr}(\mathrm{e}^{\log \tau - I - L^*(\lambda)}) - \left\langle \lambda, \bar{f} \right\rangle. \end{aligned} \tag{34}$$

The existence of such a λ° is proved in Section IV-C. Moreover, we suggest how to efficiently compute it.

Remark 4.1: When $\mu = 0$, $\mathcal{S}$ only contains singular matrices. Instead of considering the reduced problem as we did, one could try to consider Problem 3 with relaxed constraint $\rho \geq 0$. In this situation the *Slater's condition* [36, 5.2.3], however, does not hold because $\mathcal{S}$ does not contain positive definite matrices. Hence, we cannot conclude that $\rho(\lambda^\circ)$ is the desired solution of the primal problem.

C. Dual Problem: Existence and Uniqueness of the Solution

The dual problem consists in maximizing (34) over $\mathbb{R}^m$ which is equivalent to minimize

$$J(\lambda) = \mathrm{tr}(\mathrm{e}^{\log \tau - I - L^*(\lambda)}) + \left\langle \lambda, \bar{f} \right\rangle. \tag{35}$$

This functional will be referred to as *dual function* throughout this Section. Before to prove the existence of λ° which minimizes J we need to introduce the following technical results.

First of all, note that $\left\langle \lambda^\perp, \bar{f} \right\rangle = 0$ for each $\lambda^\perp \in [\mathrm{Range}\, L]^\perp$. In fact, if $\lambda^\perp \in [\mathrm{Range}\, L]^\perp = \ker L^*$, then $L^*(\lambda^\perp) = 0$. Since $\mathcal{S} \cap \mathcal{H}_{n,+} \neq \emptyset$, there exists $\rho_f \in \mathcal{H}_{n,+}$ such that $\bar{f} = L(\rho_f)$. Thus,

$$\left\langle \lambda^\perp, \bar{f} \right\rangle = \left\langle \lambda^\perp, L(\rho_f) \right\rangle = \left\langle L^*(\lambda^\perp), \rho_f \right\rangle = \mathrm{tr}(L^*(\lambda^\perp)\rho_f) = 0. \tag{36}$$

We conclude that $\lambda^\perp$ does not affect J, i.e.

$$J(\lambda + \lambda^\perp) = J(\lambda), \quad \forall\ \lambda^\perp \in [\mathrm{Range}\, L]^\perp. \tag{37}$$

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/TIT.2013.2286087, IEEE Transactions on Information Theory

We may therefore restrict the search of the minimum point for J over $\mathrm{Range}\, L$.

Proposition 4.1: J is strictly convex over $\mathrm{Range}\, L$.

Proof: Since J is the opposite of $\mathcal{L}(\rho(\lambda), \lambda)$, it is convex over $\mathbb{R}^m$. The first and the second variation of $J(\lambda)$ in direction $\delta\lambda \in \mathbb{R}^m$ are:

$$\begin{aligned}\delta J(\lambda;\delta\lambda) &= -\operatorname{tr}\int_0^1 (\mathrm{e}^{(1-t)(\log\tau - I - L^*(\lambda))} L^*(\delta\lambda) \\ &\qquad \cdot \mathrm{e}^{t(\log\tau - I - L^*(\lambda))})\mathrm{d}t + \left\langle \delta\lambda, \bar{f}\right\rangle \\ &= -\operatorname{tr}\int_0^1 \mathrm{e}^{\log\tau - I - L^*(\lambda)}\mathrm{d}t L^*(\delta\lambda) + \left\langle \delta\lambda, \bar{f}\right\rangle \\ &= -\operatorname{tr}(\mathrm{e}^{\log\tau - I - L^*(\lambda)} L^*(\delta\lambda)) + \left\langle \delta\lambda, \bar{f}\right\rangle \quad (38)\end{aligned}$$

$$\begin{aligned}\delta^2 J(\lambda;\delta\lambda) &= \operatorname{tr}[\int_0^1 \mathrm{e}^{(1-t)(\log\tau - I - L^*(\lambda))} L^*(\delta\lambda) \\ &\qquad \cdot \mathrm{e}^{t(\log\tau - I - L^*(\lambda))} L^*(\delta\lambda)\mathrm{d}t]. \quad (39)\end{aligned}$$

Here, we exploited the expression for the differential of the matrix exponential (see [21, Appendix IA]). Define

$$Q_t = \mathrm{e}^{t(\log\tau - I - L^*(\lambda))} \quad (40)$$

which is positive definite for each $t \in \mathbb{R}$. Thus,

$$\begin{aligned}\delta^2 J(\lambda;\delta\lambda) &= \int_0^1 \operatorname{tr}(Q_{1-t} L^*(\delta\lambda) Q_t L^*(\delta\lambda))\mathrm{d}t \\ &= \int_0^1 \operatorname{tr}(Q_t^{\frac{1}{2}} L^*(\delta\lambda) Q_{1-t} L^*(\delta\lambda) Q_t^{\frac{1}{2}})\mathrm{d}t \geq 0. \quad (41)\end{aligned}$$

Assume now that $\delta\lambda \in \mathrm{Range}\, L$. If $\delta^2 J(\lambda;\delta\lambda) = 0$, then $\operatorname{tr}(Q_t^{\frac{1}{2}} L^*(\delta\lambda) Q_{1-t} L^*(\delta\lambda) Q_t^{\frac{1}{2}}) = 0$. Since $Q_t > 0$ for each $t \in \mathbb{R}$, it follows that $L^*(\delta\lambda) = 0$. Since $\delta\lambda \in \mathrm{Range}\, L$, we get $\delta\lambda = 0$. We conclude that $\delta^2 J(\lambda;\delta\lambda) > 0$, for each $\delta\lambda \neq 0$, i.e. the statement holds. ∎

In the light of the previous result, the dual problem admits at most one solution, say λ°, over $\mathrm{Range}\, L$. If such a λ° does exist, then $\delta J(\lambda;\delta\lambda) = 0 \;\forall\, \delta\lambda \in \mathrm{Range}\, L$ which is equivalent to $-L(\mathrm{e}^{\log\tau - I - L^*(\lambda^\circ)}) + \bar{f} = 0$. It means that $\rho(\lambda^\circ)$ satisfies constraints in (8) and it is therefore the unique solution to Problem 5. It remains to show that such a λ° does exist.

Proposition 4.2: J admits a minimum point over $\mathrm{Range}\, L$.

Proof: We have to show that the continuous function J takes minimum value over $\mathrm{Range}\, L$. Observing that $J(0) = \frac{1}{e}\operatorname{tr}(\tau)$, we can restrict our search over the closed set

$$\mathcal{M} := \{\lambda \in \mathbb{R}^m \mid J(\lambda) \leq J(0)\} \cap \mathrm{Range}\, L.$$

We shall show that $\mathcal{M}$ is bounded. To this aim, consider a sequence $\{\lambda_i\}_{i\in\mathbb{N}}$, $\lambda_i \in \mathrm{Range}\, L$ such that $\|\lambda_i\| \to \infty$. It is therefore sufficient to show that $J(\lambda_i) \to \infty$, as $i \to \infty$. First of all note that the minimum singular value α of L^* restricted to $\mathrm{Range}\, L = [\ker L^*]^\perp$ is strictly positive, accordingly

$$\|L^*(\lambda_i)\| \geq \alpha\|\lambda_i\| \to \infty. \quad (42)$$

This means that $L^*(\lambda_i)$, which is an Hermitian matrix, has at least one eigenvalue β_i such that $|\beta_i|$ approach infinity. If $\beta_i \to -\infty$, then the first term of J tends to infinity and dominates the second one, accordingly $J(\lambda_i) \to \infty$. In the remaining possible case no eigenvalue of $L(\lambda_i)$ approaches $-\infty$ and $\beta_i \to \infty$. Thus, $L^*(\lambda_i) \geq MI$ where $M \in \mathbb{R}$ is a finite constant and the first term of J takes a finite value. Since $\mathcal{S} \cap \mathcal{H}_{n,+}$ is not empty, there exists $\rho_f \in \mathcal{H}_{n,+}$ such that $\bar{f} = L(\rho_f)$ and

$$\left\langle \lambda_i, \bar{f}\right\rangle = \langle L^*(\lambda_i), \rho_f\rangle = \operatorname{tr}(\rho_f^{\frac{1}{2}} L^*(\lambda_i)\rho_f^{\frac{1}{2}}) \geq M, \quad (43)$$

where we exploited the fact that $\operatorname{tr}(\rho_f) = 1$. This means that $\left\langle \lambda_i, \bar{f}\right\rangle$ cannot approach $-\infty$. Finally, $\|\rho_f^{\frac{1}{2}} L^*(\lambda_i)\rho_f^{\frac{1}{2}}\| \to \infty$, because $\rho_f > 0$. It follows that $\rho_f^{\frac{1}{2}} L^*(\lambda_i)\rho_f^{\frac{1}{2}}$, and hence $L^*(\Lambda_i)$, have at least one eigenvalue tending to ∞. Accordingly $J(\lambda_i) \to \infty$ as $i \to \infty$. We conclude that $\mathcal{M}$ is bounded and accordingly compact. By Weierstrass' theorem, J admits minimum point over $\mathcal{M}$. ∎

Finally, λ° may be computed by *e.g.* employing a *Newton algorithm with backtracking*, see Section VI in [37], which globally converges.

V. Conclusions

The proposed set of analytic results and algorithms provides a general method to find the *exact minimum relative entropy estimate of a quantum state under general assumptions*, that is, without requiring the constraints to include a full-rank solution to begin with. A general solution to the problem was missing for quantum estimation, and our feasibility analysis is of interest for the classical case as well. Summarizing, we have proposed: (i) A numerical method to decide the feasibility of Problem 1, which is of interest on his own, and compute the minimum necessary relaxation of the constraints whenever necessary to obtain at least one solution; (ii) As a side product, we are able to determine whether there is at least a full-rank solution. When this is not the case, we proved there exists, and devised a way to determine, the maximal common kernel of all the solution; (iii) We extended Georgiou's approach to maximum entropy estimation to our setting, and analyzed in depth both the primal and the dual optimization problems. The general form of a full-rank solution of the reduced problem, namely the one we obtain by removing the common kernel, is given explicitly and depends on the solution of the dual problem. The latter is proven to be a convex optimization problem with a unique solution, λ°, which can be obtained by standard numerical algorithms.

We also show that λ° is continuous with respect to the data set $\bar{f}$ (see Appendix B). Since, in view of (30), $\rho(\lambda^\circ)$ is continuous with respect to λ°, we can conclude that the solution to Problem 5 is continuous with respect to the data $\bar{f}_1 \ldots \bar{f}_m$. This is of course a desirable property, and ensures that for a increasing accuracy of the estimates, the computed solution will converge to the one corresponding to the exact statistics. A way to estimate the errors due to the finite accuracy of the data is also provided in the Appendix.

Possible extensions of the present framework include applications to state reconstruction from local marginals and its connection to entanglement generation [31], [32], as well as comparison with the existing approximate results [23] in physically meaningful situation. Lastly, the advantage offered

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/TIT.2013.2286087, IEEE Transactions on Information Theory

by the introduction of an *a priori* state in the estimation problem can be exploited to devise recursive algorithms, that update existing estimate in an optimal way relying only on partial data.

APPENDIX

A. Numerical solution for the feasibility problem

We propose a Newton-type algorithm with logarithmic barrier for numerically find one solution $\underline{v}^\circ$ to Problem 4. Using the same approach of [26, Section 4], we resort to the approximate problem

$$\min_{\underline{v}\in\mathrm{int}(\mathcal{I})} g_q(\underline{v}) \tag{44}$$

where $q>0$, and

$$g_q(\underline{v}) := qc^T\underline{v} - \log\det(H(\underline{v})). \tag{45}$$

Recall that we defined

$$\mathcal{I} = \{\underline{v} \mid \ H(\underline{v}) \geq 0\},$$

and notice that g_q is continuous and strictly convex over the set $\mathrm{int}(\mathcal{I})$. Moreover $\lim_{\underline{v}\to\partial\mathcal{I}} g_q(\underline{v}) = +\infty$ and the component v_{n^2-m+1} can be restricted to belong to a closed and bounded interval. Accordingly the approximated problem admits a unique solution, denoted by $\hat{\underline{v}}^q$, which is numerically computed by employing the Newton algorithm with backtracking, [36, 9.5]. Here, we can choose as initial guess the vector $\underline{\hat{v}}_0^q = \left[\begin{array}{ccccc} 0 & \dots & 0 & \sqrt{n}(-\lambda_{\min}(\tilde{\rho}_0)+1 & \end{array}\right]^T \in \mathrm{int}(\mathcal{I})$. Concerning the l-th Newton step, we have to solve

$$\Delta\underline{\hat{v}}_l^q = -\mathrm{H}^{-1}_{\underline{\hat{v}}_l^q}\nabla G_{\underline{\hat{v}}_l^q} \tag{46}$$

where

$$\nabla G_{\underline{\hat{v}}_l^q} = qc - \left[\begin{array}{c} \mathrm{tr}(H(\underline{\hat{v}}_l^q)^{-1}Y_1) \\ \vdots \\ \mathrm{tr}(H(\underline{\hat{v}}_l^q)^{-1}Y_{n^2-m}) \\ \mathrm{tr}(H(\underline{\hat{v}}_l^q)^{-1}X_1) \end{array}\right]$$

$$\mathrm{H}_{\underline{\hat{v}}_l^q} = \tag{47}$$

$$\left[\begin{array}{cc} \mathrm{tr}(H(\underline{\hat{v}}_l^q)^{-1}Y_1H(\underline{\hat{v}}_l^q)^{-1}Y_1) & \mathrm{tr}(H(\underline{\hat{v}}_l^q)^{-1}Y_1H(\underline{\hat{v}}_l^q)^{-1}Y_2) \ \dots \\ \mathrm{tr}(H(\underline{\hat{v}}_l^q)^{-1}Y_2H(\underline{\hat{v}}_l^q)^{-1}Y_1) & \mathrm{tr}(H(\underline{\hat{v}}_l^q)^{-1}Y_2H(\underline{\hat{v}}_l^q)^{-1}Y_2) \\ \vdots & \ddots \end{array}\right] \tag{48}$$

are the gradient and the Hessian computed at $\underline{\hat{v}}_l^q$, respectively. Note that, it is not difficult to prove that

1) The sequence $\{\underline{v}_l^q\}_{l\geq 0}$ generated by the algorithm is contained in the compact set $\mathcal{N} = \left\{\underline{v}\in\mathrm{int}(\mathcal{I}) \mid \ -\frac{1}{\sqrt{n}} \leq c^T\underline{v} \leq c^T\underline{v}_0^q\right\}$;
2) g_q is twice differentiable and strongly convex on $\mathcal{N}$;
3) The Hessian of g_q is *Lipschitz* continuous on $\mathcal{N}$.

Accordingly, the Newton algorithm with backtracking globally converges, [36, 9.5.3]. Moreover, the rate of convergence is quadratic in the last stage. Finally, note that the found solution $\underline{\hat{v}}^q$ satisfies the following inequalities, [36, p. 566],

$$c^T\underline{v}^\circ \leq c^T\underline{v}^q \leq c^T\underline{v}^\circ + \frac{n}{q} \tag{49}$$

where $\underline{v}^\circ$ is a solution to Problem 4 and $\frac{n}{q}$ the accuracy of $c^T\underline{v}^q$ with respect to the optimal value $c^T\underline{v}^\circ$.

This method works well only setting a moderate accuracy. To improve the accuracy, we can iterate the above Newton algorithm and in each iteration we gradually increase q in order to find a solution $\underline{v}^\xi$ with a specified accuracy $\xi>0$ [36, p. 569]:

1) Set the initial conditions $q_0 > 0$ and $\underline{v}^{q_0} = \left[\begin{array}{ccccc} 0 & \dots & 0 & \sqrt{n}(-\lambda_{\min}(\tilde{\rho}_0)+1) & \end{array}\right]^T \in \mathrm{int}(\mathcal{I})$.
2) At the k-th iteration compute $\underline{v}^{q_k}\in\mathrm{int}(\mathcal{I})$ by minimizing g_{q_k} with starting point $\underline{v}^{q_{k-1}}$ by using the Newton method previously presented.
3) Set $q_{k+1}=\beta q_k$ with $\beta>1$ closer to one.
4) Repeat steps 2 and 3 until the condition $\frac{n}{q_k} < \xi$ is satisfied.

Finally, we deal with the problem to compute a solution to Problem 1 with kernel $\mathcal{K}$ when $\mu=0$. In this situation, consider the non-empty convex set

$$\mathcal{I}_* = \left\{\underline{w} \mid \tilde{\rho}_0 + \sum_{i=1}^{n^2-m} w_iY_i \geq 0\right\}, \tag{50}$$

where $\underline{w} = \left[\begin{array}{cccc} w_1 & \dots & w_{n^2-m} & \end{array}\right]^T \in \mathbb{R}^{n^2-m}$. Thus, the set of solutions to Problem 1 is

$$\mathcal{S} = \left\{\tilde{\rho}_0 + \sum_{i=1}^{n^2-m} w_iY_i \mid \underline{w}\in\mathcal{I}_*\right\}. \tag{51}$$

We wish to compute a matrix $\rho^\circ\in\mathcal{S}$ having kernel corresponding to the minimum common kernel $\mathcal{K}$. To this aim consider the following problem.

Problem 6: Pick $\underline{u}\in\mathbb{R}^{n^2-m}$ at random and solve:

$$\underline{w}^u = \arg\min_{\underline{w}\in\mathcal{I}_*} (\underline{w}-\underline{u})^T(\underline{w}-\underline{u}). \tag{52}$$

Note that:

- $\mathcal{I}_*$ is compact (i.e. closed and bounded)
- $(\underline{w}-\underline{u})^T(\underline{w}-\underline{u})$ is continuous and strictly convex on $\mathcal{I}_*$.

By Weiestrass' theorem, the above problem admits a unique solution $\underline{w}^u$.

Let us define $\rho^{\circ,u} := \tilde{\rho}_0 + \sum_{k=1}^{n^2-m} w_i^uY_i \in \mathcal{S}$ and $\rho^u := \tilde{\rho}_0 + \sum_{k=1}^{n^2-m} u_iY_i \in \mathcal{S}$. It is easy to see that Problem 6 can be rewritten in terms of the matrix $\rho^{\circ,u}$ as follows:

$$\rho^{\circ,u} = \arg\min_{\rho\in\mathcal{S}} \|\rho^u-\rho\|_F^2, \tag{53}$$

where $\|\cdot\|_F$ denotes the *Frobenius* matrix norm. Thus, $\rho^{\circ,u}$ is the closest matrix in $\mathcal{S}$ (with respect to the *Frobenius* norm) to the matrix ρ^u belonging to the hyperplane characterized by the constraints (8). Hence, randomly generating a finite sequence $\{\underline{u}_1\dots\underline{u}_l\}$ of elements in $\mathbb{R}^{n^2-m}$ we obtain a subset $\mathcal{U} := \{\rho^{\circ,u_1}\dots\rho^{\circ,u_l}\}$ contained in $\mathcal{S}$. Construct the convex combination $\bar{\rho} := \frac{1}{l}\sum_{j=1}^{l}\rho^{\circ,u_j}$. Then $\bar{\rho}$ has minimal kernel if either one of the ρ°,u_j} belongs to the interior of $\mathcal{S}$, or ρ°,u_j} belong to different faces of $\mathcal{S}$, which is a compact convex set. By the randomized construction, the probability of remaining on the boundary of $\mathcal{S}$ becomes small as l grows.

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/TIT.2013.2286087, IEEE Transactions on Information Theory

Concerning the computation of $\underline{w}^u$, we can resort a Newton-type algorithm with logarithmic barrier named *exterior-point method*, [38, Chapter 4].

B. Continuity of λ° with respect to $\bar{f}$

We show that the solution λ° is continuous with respect to the data set $\bar{f}$. To this aim we take into account the following result, see [39, Theorem 3.1].

Theorem A.1: Let A be an open and convex subset of a finite-dimensional euclidean space V. Let $h : A \to \mathbb{R}$ be a strictly convex function, and suppose that a minimum point $\bar{x}$ of h exists. Then, for all $\varepsilon > 0$, there exists $\delta > 0$ such that, for $p \in \mathbb{R}^n$, $\|p\| < \delta$, the function $h_p : A \to \mathbb{R}$ defined as

$$h_p(x) := h(x) - \langle p, x \rangle \tag{54}$$

admits a unique minimum point $\bar{x}_p$, and moreover

$$\|\bar{x}_p - \bar{x}\| < \varepsilon. \tag{55}$$

In order to exploit this result, consider

$$J(\lambda, \bar{f}) = \mathrm{tr}(\mathrm{e}^{\log \tau - I - L^*(\lambda)}) + \left\langle \lambda, \bar{f} \right\rangle \tag{56}$$

where we make explicit the dependence of J upon $\bar{f}$. Then, the unique minimum point is

$$\lambda(\bar{f}) = \arg \min_{\lambda \in \mathbb{R}^m} J(\lambda, \bar{f}). \tag{57}$$

Let $\delta f \in \mathbb{R}^m$ be a perturbation of $\bar{f}$. We have $J(\lambda, \bar{f} + \delta f) = J(\lambda, \bar{f}) + \lambda^T \delta f$. Applying the previous theorem, where δf is $-p$, we have: $\forall\ \varepsilon > 0\ \exists\ \delta > 0$ s.t. if $\|\delta f\| < \delta$ then $J(\lambda, \bar{f} + \delta f)$ admits a unique minimum point

$$\lambda(\bar{f} + \delta f) = \arg \min_{\lambda \in \mathbb{R}^m} J(\lambda, \bar{f} + \delta f) \tag{58}$$

and

$$\|\lambda(\bar{f} + \delta f) - \lambda(\bar{f})\| < \varepsilon. \tag{59}$$

Accordingly, the map $\bar{f} \mapsto \lambda(\bar{f})$ is continuous.

Remark: By exploiting the ideas of the proof of Theorem 3.1 in [39], we can easily obtain an estimate and a lower bound of the constant δ as a function of ε. In fact, following [39], we can pick

$$\delta = \frac{J(\lambda^\circ(\bar{f}) + \varepsilon \delta\mu^\circ) - J(\lambda^\circ(\bar{f}))}{\varepsilon}$$

where $\delta\mu^\circ$ is a perturbation that minimizes $J(\lambda^\circ(\bar{f}) + \varepsilon\delta\mu)$ over the set $\{\delta\mu :\ \|\delta\mu\| = 1\}$. Let us consider a second order Taylor expansion of $J(\lambda^\circ(\bar{f}) + \varepsilon\delta\mu^\circ)$:

$$J(\lambda^\circ(\bar{f}) + \varepsilon\delta\mu^\circ) = J(\lambda^\circ(\bar{f})) + \varepsilon^2 \cdot \frac{1}{2!}\delta^2 J(\lambda^\circ(\bar{f}) + \delta\lambda, \delta\mu^\circ)$$

for a certain $\|\delta\lambda\| \le \varepsilon$. Here $\delta^2 J$ is the second variation of J. We immediately get that

$$\delta = \varepsilon \cdot \frac{1}{2}\delta^2 J(\lambda^\circ(\bar{f}) + \delta\lambda, \delta\mu^\circ)$$

with $\|\delta\mu^\circ\| = 1$. Therefore we have

$$\delta \ge \frac{1}{2} \min_{\|\delta\mu\|=1,\ \|\delta\lambda\| \le \varepsilon} \varepsilon \cdot \delta^2 J(\lambda^\circ(\bar{f}) + \delta\lambda, \delta\mu).$$

If we exploit the explicit expression (41) for the second variation $\delta^2 J$, we can directly compute a lower bound for $\delta^2 J(\lambda^\circ(\bar{f}) + \delta\lambda, \delta\mu)$. In fact,

$$\begin{aligned} &\frac{1}{2} \min_{\|\delta\mu\|=1,\ \|\delta\lambda\| \le \varepsilon} \delta^2 J(\lambda^\circ(\bar{f}) + \delta\lambda, \delta\mu) \\ &\ge K := \frac{1}{2}\sigma_m^2 \min_{\|\delta\mu\|=1} \mathrm{tr}[(L^*(\delta\mu))^2] \end{aligned}$$

where σ_m^2 is the minimum eigenvalue of Q_t (as defined in (40)) over $t \in [0, 1]$ and $\lambda = \lambda^\circ(\bar{f}) + \delta\lambda$, with $\|\delta\lambda\| \le \varepsilon$. Let us now focus on the right hand side of the inequality. The value of σ_m^2 is very easy to compute: first notice that this minimum is either attained for $t = 0$ (in which case it equals 1) or for $t = 1$. Next, since the matrix Q_t is the exponential of an affine function of λ, in order to compute its minimum eigenvalue is it sufficient to compute the minimum eigenvalue of $A(\lambda) := \log \tau - I - L^*(\lambda)$ as λ varies in $\lambda = \lambda^\circ(\bar{f}) + \delta\lambda$, with $\|\delta\lambda\| \le \varepsilon$. Similarly, since $L^*(\delta\mu)$ is a linear function of $\delta\mu$, we can easily compute

$$\min_{\|\delta\mu\|=1} \mathrm{tr}[(L^*(\delta\mu))^2].$$

Summing up, we have derived an inequality of the form

$$\delta \ge K\varepsilon$$

where K can be explicitly computed for the specific problem. This inequality can serve a twofold purpose: on the one hand, once we fix the desired precision ε with which we would like to obtain the optimal solution, we have a bound for the precision δ up to which the $\bar{f}$ should be known. Being these obtained from repeated measurements with i.i.d. distribution, the needed number of trials can be estimated using the central limit theorem. Conversely, if due to limited statistical confidence on the data the measurements $\bar{f}$ are known up to a certain tolerance δ, we can estimate the maximum error ε on the precision with which the optimal (dual) solution $\lambda^\circ(\bar{f})$ will be computed. Accordingly, since the optimal solution $\rho(\lambda^\circ)$ of Problem 5 is a known (continuous) function of λ°, we can estimate the maximum error on the precision with which $\rho(\lambda^\circ(\bar{f}))$ will be computed.

REFERENCES

[1] R. P. Feynman, "Simulating physics with computers," *International Journal of Theoretical Physics*, vol. 21, 1982.

[2] D. Bouwmeester, A. Ekert, and A. Zeilinger, Eds., *The Physics of Quantum Information: Quantum Cryptography, Quantum Teleportation, Quantum Computation*. Springer-Verlag, 2000.

[3] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Information*. Cambridge University Press, Cambridge, 2002.

[4] D. D'Alessandro, *Introduction to Quantum Control and Dynamics*, ser. Applied Mathematics & Nonlinear Science. Chapman & Hall/CRC, 2007.

[5] H. M. Wiseman and G. J. Milburn, *Quantum Measurement and Control*. Cambridge University Press, 2009.

[6] C. Altafini and F. Ticozzi, "Modeling and control of quantum systems: An introduction," *IEEE Trans. Aut. Cont,*, vol. 57, no. 8, pp. 1898 –1917, 2012.

[7] M. Paris and J. Rehacek, Eds., *Quantum State Estimation (Lecture Notes in Physics, vol.649)*, ser. Lecture Notes in Physics. Springer, 2004, vol. 649.

[8] D. Petz, *Quantum Information Theory and Quantum Statistics*. Springer Verlag, 2008.

[9] M. Mohseni, A. T. Rezakhani, and D. A. Lidar, "Quantum-process tomography: Resource analysis of different strategies," *Phys. Rev. A*, vol. 77, p. 032322, 2008.

[10] I. Bongioanni, L. Sansoni, F. Sciarrino, G. Vallone, and P. Mataloni, "Experimental quantum process tomography of non-trace-preserving maps," *Phys. Rev. A*, vol. 82, no. 4, p. 042307, 2010.

[11] P. Villoresi, T. Jennewein, F. Tamburini, C. B. M. Aspelmeyer, R. Ursin, C. Pernechele, V. Luceri, G. Bianco, A. Zeilinger, and C. Barbieri, "Experimental verification of the feasibility of a quantum channel between space and earth," *New Journal of Physics*, vol. 10, p. 033038, 2008.

[12] E. T. Jaynes, "Information Theory and Statistical Mechanics," *Physical Review Series*, vol. 106, no. 4, pp. 620–630, 1957.

[13] ——, "Information Theory and Statistical Mechanics. II," *Physical Review Series*, vol. 108, no. 2, pp. 171–190, 1957.

[14] J. Burg, *Maximum entropy spectral analysis*, ser. Stanford Exploration project. Stanford University, 1975.

[15] D. Petz, "Entropy, von Neumann and the von Neumann entropy," in *John von Neumann and the Foundations of Quantum Physics*, M. Redei and M. Stoltzner, Eds. Kluwer, 2001.

[16] C. Helstrom, *Quantum Detection and Estimation Theory*. Academic Press, 1976.

[17] A. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory*. North-Holland, Amsterdam, 1982.

[18] V. Bužek, G. Drobný, G. Adam, R. Derka, and P. Knight, "Reconstruction of quantum states of spin systems via the jaynes principle of maximum entropy," *Journal of Modern Optics*, vol. 44, no. 11-12, pp. 2607–2627, 1997.

[19] M. Ziman, "Incomplete quantum process tomography and principle of maximal entropy," *Phys. Rev. A*, vol. 78, p. 032118, 2008.

[20] J. Shore and R. Johnson, "Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy," *IEEE Trans. Inf. Theor.*, vol. 26, no. 1, pp. 26–37, Jan. 1980.

[21] T. Georgiou, "Relative entropy and the multivariable multidimensional moment problem," *Information Theory, IEEE Transactions on*, vol. 52, no. 3, pp. 1052 –1066, 2006.

[22] V. Vedral, "The role of relative entropy in quantum information theory," *Rev. Mod. Phys.*, vol. 74, no. 1, pp. 197–234, 2002.

[23] S. Olivares and M. G. A. Paris, "Quantum estimation via the minimum kullback entropy principle," *Phys. Rev. A*, vol. 76, p. 042120, 2007.

[24] M. Pavon and F. Ticozzi, "Discrete-time classical and quantum markovian evolutions: Maximum entropy problems on path space," *J. Math. Phys.*, vol. 51, p. 042104, 2010.

[25] F. Ticozzi and M. Pavon, "On time-reversal and space-time harmonic processes for markovian quantum channels," *Quantum Information Processing*, vol. 9, no. 5, pp. 551–574, Oct. 2010.

[26] M. Zorzi, F. Ticozzi, and A. Ferrante, "On quantum channel estimation with minimal resources," *on-line preprint*, vol. arxiv.org/abs/1106.2105, 2011.

[27] ——, "Estimation of quantum channels: Identifiability and ml methods," in *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*, 2012, pp. 1674–1679.

[28] S. L. Braunstein, "Geometry of quantum inference," *Physics Letters A*, vol. 219, no. 3-4, pp. 169–174, 1996.

[29] A. Aiello, G. Puentes, D. Voigt, and J. P. Woerdman, "Maximum-likelihood estimation of mueller matrices," *Opt. Lett.*, vol. 31, no. 6, pp. 817–819, Mar 2006.

[30] J. J. Sakurai, *Modern Quantum Mechanics*. Addison-Wesley, New York, 1994.

[31] W. Hall, "Compatibility of subsystem states and convex geometry," *Phys. Rev. A*, vol. 75, p. 032102, Mar 2007.

[32] F. Ticozzi and L. Viola, "Stabilizing entangled states with quasi-local quantum dynamical semigroups," *Phil. Trans. R. Soc. A*, vol. 370, no. 1979, pp. 5259–5269, 2012.

[33] A. Jamiolkowski, "Linear transformations which preserve trace and positive semidefiniteness of operators," *Reports on Mathematical Physics*, vol. 3, no. 4, pp. 275–278, 1972.

[34] M. D. Choi, "Completely positive linear maps on complex matrices," *Linear algebra and its applications*, vol. 10, pp. 285–290, 1975.

[35] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information (Cambridge Series on Information and the Natural Sciences)*. Cambridge University Press, 2004.

[36] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.

[37] A. Ferrante, M. Pavon, and M. Zorzi, "A maximum entropy enhancement for a family of high-resolution spectral estimators," *IEEE Trans. Autom. Control*, vol. 57, no. 2, pp. 318–329, Feb. 2012.

[38] A. Fiacco and G. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*. Society for Industrial and Applied Mathematics, 1987.

[39] F. Ramponi, A. Ferrante, and M. Pavon, "On the well-posedness of multivariate spectrum approximation and convergence of high-resolution spectral estimators," *System & Control Letters*, vol. 59, no. 3-4, pp. 167–172, 2010.

Mattia Zorzi was born in Padova, Italy, in January 1985. He received the M.S. degree (with high honors) in automation engineering and the Ph.D. degree in control engineering from University of Padova, Padova, Italy, in 2009 and 2012, respectively. He is currently a post-doctoral Fellow within the Systems and Modeling group, Université de Liège, Liège, Belgium. His current research interests include graphical models, spectral estimation, robust estimation and quantum systems estimation.

Francesco Ticozzi received the "Laurea" degree in management engineering and a Ph.D. in automatic control and operations research from the University of Padova, Italy, in 2002 and 2007, respectively.

Since February 2007, he has been with the Department of Information Engineering at the University of Padova, first as a Research Associate, and then as an Assistant Professor (*Ricercatore*). During 2005–2010 he held visiting appointments at the Physics and Astronomy Dept. of Dartmouth College, Hanover, New Hampshire, where he is an Adjunct Assistant Professor since July 2011.

Dr. Ticozzi's research interests include modeling and control of quantum systems, protection of quantum information, quantum communications and information-theoretic methods in control theory.

Augusto Ferrante was born in Piove di Sacco, Italy, on August 5, 1967. He received the "Laurea" degree, *cum laude*, in Electrical Engineering in 1991 and the Ph.D. degree in Control Systems Engineering in 1995, both from the University of Padova.

He has been on the faculty of the Colleges of Engineering of the University of Udine, and of the "Politecnico di Milano". He is presently Professor in the "Department of Information Engineering" of the University of Padova.

His research interests are in the areas of linear systems, spectral estimation, optimal control and optimal filtering, quantum control, and stochastic realization.