

Article

A Maximum Entropy Approach to the Realizability of Spin Correlation Matrices

Paolo Dai Pra , Michele Pavon * and Neeraja Sahasrabudhe

Department of Mathematics, University of Padova, via Trieste 63, 35121 Padova, Italy;
E-Mails: daipra@math.unipd.it (P.D.P.); neeraja@math.unipd.it (N.S.)

* Author to whom correspondence should be addressed; E-Mail: pavon@math.unipd.it;
Tel.: +39-049-827-1341; Fax: +39-049-827-1428.

Received: 26 February 2013; in revised form: 10 June 2013 / Accepted: 15 June 2013 /
Published: 21 June 2013

Abstract: Deriving the form of the optimal solution of a maximum entropy problem, we obtain an infinite family of linear inequalities characterizing the polytope of spin correlation matrices. For $n \leq 6$, the facet description of such a polytope is provided through a minimal system of Bell-type inequalities.

Keywords: correlation matrix; spin system; maximum entropy; Bell's inequalities; moment problem

1. Introduction

Moment problems are fairly common in many areas of applied mathematics, statistics and probability, economics, engineering, physics and operations research. Historically, moment problems came into focus with Stieltjes in 1894 [1], in the context of studying the analytic behavior of continued fractions. The term “moment” was borrowed from mechanics: the moments could represent the total mass of an unknown mass density, the torque necessary to support the mass on a beam, *etc.* Over time, however, moment problems took the shape of an important field in their own right. A deep connection with convex geometry was discovered by Krein in the mid 1930s and developed by the Russian school; see, e.g., [2,3]. Another fundamental connection with the work of Caratheodory, Toeplitz, Schur, Nevanlinna and Pick on analytic interpolation was investigated in the first half of the twentieth century [4]. This led to important developments in operator theory; see, e.g., [5,6]. In more recent times, a rather impressive application and generalization of this mathematics has been developed by the

Byrnes-Georgiou-Lindquist school for signal and control engineering applications. In their approach, one seeks a multivariate, stationary stochastic process as the input of a bank of rational filters whose output covariance has been estimated. This turns into a Nevanlinna-Pick interpolation problem with a bounded degree [7,8]. The latter can be viewed as a generalized moment problem (namely, a moment problem with complexity constraints), which is advantageously cast in the frame of various convex optimization problems, often featuring entropic-type criteria. An example is provided by the covariance extension problem and its generalization; see [9–14]. These problems pose a number of theoretical and computational challenges, especially in the multivariable framework, for which we also refer the reader to [15–22]. Besides signal processing, significant applications of this theory are found in modeling and identification [23–25], H_∞ robust control [26,27], and biomedical engineering [28].

A general moment problem can be stated as follows. Suppose we are given a measurable space, (Ω, \mathcal{A}) , a family, \mathcal{F} , of measurable functions for Ω to \mathbb{R} and a corresponding family of real numbers, $\{c_f : f \in \mathcal{F}\}$. One wants to determine whether there exists a probability, P , on (Ω, \mathcal{A}) , such that:

$$E_P(f) := \int f dP = c_f$$

for every $f \in \mathcal{F}$ and, if so, characterize all probabilities having this property.

Among the various instances of this inverse problem, the covariance realization/completion problem has raised wide interest, in part because of its important applications in mathematical statistics [29] and in theoretical engineering [30]. It is well known that an $n \times n$ matrix is the covariance matrix of some \mathbb{R}^n -valued random vector if and only if it belongs to the convex cone of symmetric and positive semidefinite matrices. A relevant problem for applications considers the situation in which only some entries of the covariance matrix are given, for example, those that have been estimated from the data. In this context, one aims at characterizing all possible completions of the partially given covariance matrix or completions that possess certain desirable properties; see, e.g., [29,31–37] and references therein. Another more theoretical problem investigates the geometry of correlation matrices, namely, covariances of standardized random variables. Clearly, correlation matrices form a compact, convex subset of the vector space of symmetric matrices of dimension, n , since the latter is determined by a family of linear inequalities, namely, the positivity constraints. A natural question is to determine the extreme points of this convex set. This problem was solved by Parthasarathy in [38], who could parametrize the uncountable family of extremals, which turn out to be all singular.

The geometry of correlation matrices may change dramatically if one adds constraints on the values of the random vector realizing the covariance matrix. In this paper, we consider the case in which the components of the random vector are required to be $\{-1, 1\}$ -valued. We call spin systems random vectors of this type or, by abuse of language, their distributions. Although spin systems have been extensively studied in statistical mechanics and probability, various questions are still open concerning their covariance matrices. In [39], J. C. Gupta proved that covariance matrices of a system of n spins form a polytope and exhibited its 2^{n-1} extremals. Apparently, his result is contained in some form in Pitowsky's previous work [40] and in even earlier, but not easily accessible, work by Assouad (1980) and Assouad and Deza (1982); cf. [41] (Section 5.3), for more information.

A more delicate problem is to characterize covariance matrices of spin systems by a system of linear inequalities. This problem was tackled in [40] (see, however [41] (p.54) for a thorough description of

preceding contributions on “correlation polytopes” coming from such different fields as mathematical physics, quantum logic and analysis). There, the dual description, in the sense of linear programming, was considered, and the high complexity of the problem of determining the extremals of this dual, *i.e.*, the facets of the polytope, was discussed. Moreover, in [42], it has been shown that this dual is generated by the Bell’s inequalities (see [43] for an overview of the role played by these inequalities in quantum mechanics) for $n = 3, 4$, but not for $n \geq 5$. Finally, we mention the paper [44], where the problem of realizability of correlations has been extended to the more general setting of random points in a topological space.

The aim of this paper is two-fold.

- We derive an infinite family of linear inequalities characterizing covariances of spin systems, via the solution of a maximum entropy problem. Besides its intrinsic interest, this method has the advantage of describing, in terms of certain Lagrangian multipliers, an explicit probability realizing the covariances, whenever they are realizable. The search for the Lagrange multipliers is an interesting computational problem, which will be addressed in a forthcoming paper.
- Via a computer-aided proof, we determine the facets of the polytope of covariance matrices of spin systems for $n \leq 6$. In particular, we show that for these values of n , Bell’s inequalities are actually facets of the polytope, but generate the whole polytope only for $n = 3, 4$. For $n = 5$ and 6 , the remaining facets are given by suitable generalizations of Bell’s inequalities. Although the problem is computationally feasible also for some larger values of n , the number of extremal inequalities increases dramatically, and we have not been able to describe them synthetically. We mention the fact that the case $n = 3$ is peculiar, since it is the only case in which the polytope is a simplex. A more detailed description of this case is contained in the note [45]. Our work here inevitably overlaps with some previous research on linear descriptions of polytopes in combinatorial geometry, such as [46]; see also (Section 30.6 in [41]) and, in particular, the footnote on p. 503 of the latter reference (the book [41] by M. Deza and M. Laurent is a general, comprehensive reference on discrete geometry). We remark that our arguments go through even when the covariance matrix is only partially given, a case important for applications, but typically not considered in the discrete geometry literature.

Summing up, we obtain necessary and sufficient conditions for the existence of a covariance completion, as well as a “canonical” (maximum entropy) probability realizing the given covariances.

2. Spin Systems and Spin Correlation Matrices

Let us define a spin system first. Let $\Omega_n = \{-1, 1\}^n$ be the space of length- n sequences, which are denoted by $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_n)$, where $\sigma_i \in \{-1, 1\}$. Define the spin random variables, $\xi_i : \Omega_n \rightarrow \{-1, 1\}$, for $1 \leq i \leq n$ as $\xi_i(\sigma) = \sigma_i$. For a probability, P , on Ω_n , we denote by E_P the expectation with respect to P . The finite probability space (Ω_n, P) is called the spin system. As in [39], for simplicity, we only consider symmetric probabilities, *i.e.*, those for which $E_P(\xi_i) = 0$ for all $1 \leq i \leq n$. Note that, in this case, the covariance matrix, $C = E_P(\xi_i \xi_j)_{i,j=1}^n$, has all diagonal elements equal to 1. We will refer to this matrix as the spin-correlation matrix associated to P .

Suppose that we are given the spin-spin correlations, c_{ij} 's, and look for a probability, P , on Ω_n , such that $c_{ij} = E_P(\xi_i \xi_j)$ for every $1 \leq i, j \leq n$ or for all pairs (i, j) for which c_{ij} is given. We consider the following questions:

- Under what conditions does a distribution with those correlations exist?
- If one such distribution exists, that is, if the given correlations are realizable, then how does one characterize the maximum entropy probability measure?

The spin correlation/covariance matrices form a convex polytope whose description in terms of vertices is known. Let us denote the convex polytope of the spin correlation matrices of the order, n , by Cov_n . J. C. Gupta proved that Cov_n is the convex hull of 2^{n-1} matrices (it turns out that these matrices are exactly the extremal vertices of this polytope) that can be found explicitly.

Theorem 1 (J. C. Gupta, 1999): *The class of realizable correlation matrices of n spin variables is given by:*

$$\text{Cov}_n = \text{ConvexHull}\{\Sigma^S : S \in \mathcal{S}\}$$

where Σ^S and \mathcal{S} are defined as follows:

$$\mathcal{S} = \{S \subset \{1, 2, \dots, n\} : 1 \in S\}$$

and

$$\Sigma^S = ((c_{ij}^S))$$

where $c_{ii}^S = 1$ for all i and $c_{ij}^S = (-1)^{|S \cap \{i,j\}|}$ for $i \neq j$.

These Σ^S 's are rank-1 matrices obtained by considering probability measures on Ω_n supported at two points (configurations), such that each of these two configurations have probability, $1/2$. The proof of above theorem can be found in [39].

It is interesting to note that the description of extremals of spin correlation matrices is rather simple when compared with the description of extremals of the convex set of correlation matrices in general (see [38]). As mentioned in the introduction, it is more difficult to obtain the dual representation in terms of linear inequalities. One simple observation is that every spin correlation matrix, $C = (c_{ij})$, must satisfy the following Bell's inequalities: for every $\varepsilon \in \Omega_n$ and $1 \leq i < j < k \leq n$:

$$1 + \varepsilon_i \varepsilon_j c_{ij} + \varepsilon_j \varepsilon_k c_{jk} + \varepsilon_k \varepsilon_i c_{ki} \geq 0 \tag{1}$$

The necessity of these inequalities is easy to show:

$$\begin{aligned} \varepsilon_i \varepsilon_j c_{ij} + \varepsilon_j \varepsilon_k c_{jk} + \varepsilon_k \varepsilon_i c_{ki} &= E_P [\varepsilon_i \varepsilon_j \xi_i \xi_j + \varepsilon_j \varepsilon_k \xi_j \xi_k + \varepsilon_k \varepsilon_i \xi_k \xi_i] \\ &= \frac{1}{2} E_P [(\varepsilon_i \xi_i + \varepsilon_j \xi_j + \varepsilon_k \xi_k)^2] - \frac{3}{2} \geq -1 \end{aligned}$$

where, in the last step, we have observed that $(\sigma_i + \sigma_j + \sigma_k)^2 \geq 1$ for every $\sigma \in \Omega_n$. One immediate consequence is that not all positive matrices, with diagonal elements equal to 1, are spin correlation matrices.

Example 1 Consider a symmetric matrix:

$$C = \begin{pmatrix} 1 & c_1 & c_2 \\ c_1 & 1 & c_1 \\ c_2 & c_1 & 1 \end{pmatrix}$$

with $-1 \leq c_1, c_2 \leq 1$. Then, the condition for positive-definiteness is:

$$1 - 2c_1^2 + c_2 \geq 0$$

and the Bell’s inequalities are given by:

$$1 \pm 2c_1 + c_2 \geq 0$$

For instance, for $c_1 \in (-\frac{1}{\sqrt{2}}, -\frac{1}{2}) \cup (\frac{1}{2}, \frac{1}{\sqrt{2}})$ and $c_2 = 0$, we get a matrix, C , that is symmetric and positive semi-definite—hence, a correlation matrix—but it does not satisfy the Bell’s inequalities, so it can’t be a spin correlation matrix.

3. The Dual Representation for $n \leq 6$

We have seen that the spin correlation matrices form a convex polytope with extreme points given by Theorem 1. Every convex polytope has two representations: one as the convex hull of finitely many extreme points (known as the V -representation) and another in terms of the inequalities defining the faces of the polytope (known as the H -representation). These inequalities provide necessary and sufficient conditions for a point to lie inside the convex hull. Thus, finding necessary and sufficient conditions for a matrix, M , to lie in Cov_n is equivalent to finding the H -representation of Cov_n . The problem of obtaining the H -representation from the V -representation is called the facet enumeration problem, while the dual one is called the vertex enumeration or the convex hull problem. These are well known problems in the theory of linear programming.

The program, `cdd+` (`cdd`, respectively), is a C++ (ANSI C) program that performs both tasks. Given the equations of faces of the polytope, it returns the set of vertices and extreme rays and *vice versa* [47]. This program is a computer implementation of the double description method (see, for instance, [48]). This program works with integer arithmetics; in particular, when data in the input are integers, it does not make any rounding.

We executed the `cdd+` program to find the necessary and sufficient condition for $3 \leq n \leq 6$. We know the extremals in each case from Theorem 1. We summarize below the results obtained. We remark that the facets of this correlation polytope have been previously computed for $n \leq 8$ (Section 30.6 in [41]). Our point here is to connect these facets with Bell’s inequalities and their generalizations (see Section 4).

3.1. Cases $n = 3, 4$

These are the simplest cases, already covered in [42]. The program returns exactly the Bell’s inequalities in Equation (1). In particular, the following nontrivial facts follow:

- Bell’s inequalities imply positivity of the matrix;

- Bell’s inequalities correspond to the facets of the polytope of spin correlation matrices in dimension three and four; in particular, they provide the “minimal” description in terms of linear inequalities.

3.2. Case $n = 5$

The polytope of spin-correlation matrices has 56 facets. Forty of these are given by the Bell’s inequalities, corresponding to $\binom{5}{3} = 10$ choices of three indexes and 2^2 for $\varepsilon \in \{-1, 1\}^3$ (modulo change of sign).

The remaining 16 facets correspond to the following inequalities: for every $\varepsilon \in \{-1, 1\}^5$, the modulo sign:

$$2 + \sum_{1 \leq i < j \leq 5} \varepsilon_i \varepsilon_j c_{ij} \geq 0 \tag{2}$$

3.3. Case $n = 6$

There are 368 facets. We can group the corresponding inequalities into three groups.

1. We have the $\binom{6}{3} 2^2 = 80$ Bell’s inequalities.
2. For $T \subseteq \{1, 2, 3, 4, 5, 6\}$ with $|T| = 5$ and $\varepsilon \in \{-1, 1\}^T$, we consider the inequality analogous to Equation (2):

$$2 + \sum_{i < j; i, j \in T} \varepsilon_i \varepsilon_j c_{ij} \geq 0$$

There are $\binom{6}{5} 2^4 = 96$ inequalities of this sort.

3. For $T \subseteq \{1, 2, 3, 4, 5, 6\}$ with $|T| = 5$ and $\varepsilon \in \{-1, 1\}^T$, letting j_T be the only element of T^c , consider the inequalities:

$$4 + \sum_{i < j; i, j \in T} \varepsilon_i \varepsilon_j c_{ij} + 2 \sum_{i \in T} \varepsilon_i \varepsilon_{j_T} c_{ij_T} \geq 0$$

There are $\binom{6}{5} 2^5 = 192$ such inequalities.

We will see in the next section that inequalities of the types above hold for spin-correlation matrices also in higher dimensions, where, however, facets of different types appear.

4. Maximum Entropy Measure for Spin Systems

4.1. Maximum Entropy Method

Our aim now is to find an explicit measure that realizes the given covariances. One of the most natural and popular approach in these kind of problems is to use the maximum entropy method. The rationale underline this approach has been discussed over the years by various “deep thinkers” such as Jaynes [49–51] (physics), Dempster [29] (statistics) and Csiszár [52] (information theory). We refer the reader to these references for full motivation of this approach.

We want to find a probability measure that realizes the given covariances and that also maximizes the entropy of the system. In other words, we want to solve the following optimization problem:

Maximize the entropy:

$$\mathcal{S}(P) := - \sum_{\sigma} P(\sigma) \ln P(\sigma)$$

over

$$P : \Omega_n \rightarrow [0, \infty)$$

subject to:

$$\begin{aligned} \sum_{\sigma} \sigma_h \sigma_k P(\sigma) &= c_{hk} \\ \sum_{\sigma} P(\sigma) &= 1 \end{aligned}$$

Consider the *Lagrangian function*:

$$\mathcal{L}(P) = \mathcal{S}(P) + \sum_{h,k} \lambda_{hk} \left(c_{hk} - \sum_{\sigma} \sigma_h \sigma_k P(\sigma) \right) + \mu \left(\sum_{\sigma} P(\sigma) - 1 \right)$$

Notice that $\mathcal{L}(P)$ coincides with $\mathcal{S}(P)$ on the set of P satisfying the constraints:

$$M = \left\{ P(\sigma) \mid P(\sigma) \geq 0, \sum_{\sigma} P(\sigma) = 1, \sum_{\sigma} \sigma_h \sigma_k P(\sigma) = c_{hk} \right\}$$

Here, $\mu \in \mathbb{R}$ and the $n \times n$ matrix $\Lambda = (\lambda_{hk})$ are the *Lagrange multipliers*. $\mathcal{L}(P)$ is a strictly concave function of P on the convex cone, \mathcal{P} , of positive measures on Ω_n . Thus, if we can find an internal point, $P^* \in \mathcal{P}$, such that:

$$\mathcal{L}'(P^*; \delta P) = \lim_{\epsilon \rightarrow 0} \frac{\mathcal{L}(P^* + \epsilon \delta P) - \mathcal{L}(P^*)}{\epsilon} = 0$$

for all $\delta P : \Omega_n \rightarrow \mathbb{R}$, then, necessarily, P^* is the unique maximum point for \mathcal{L} over \mathcal{P} . Since:

$$\mathcal{L}'(P^*; \delta P) = \sum_{\sigma} \left[-\log P^*(\sigma) - \sum_{h,k} \lambda_{hk} \sigma_h \sigma_k + \mu - 1 \right] \delta P(\sigma)$$

we get the optimality condition:

$$-\log P^*(\sigma) - \sum_{h,k} \lambda_{hk} \sigma_h \sigma_k + \mu - 1 \equiv 0$$

Namely, P^* has the form:

$$P^*(\sigma) = \frac{1}{Z} \exp \left\{ - \sum_{h,k} \lambda_{hk} \sigma_h \sigma_k \right\}, \quad \sigma \in \Omega_n \tag{3}$$

where $Z = \exp \{1 - \mu\}$. Such a P^* is, in fact, an internal point of \mathcal{P} . Note that any probability of this form is such that $P^*(\sigma) = P^*(-\sigma)$. In particular, this implies that each spin has a mean of zero with respect to P^* .

Also note that this last formula simply specifies a class of probability measures on Ω_n , parametrized by the matrix of Lagrange multipliers, $\Lambda = (\lambda_{ij})$. It remains to establish whether the given correlations are realized by any such probability and, if so, to determine the corresponding values of the multipliers. To this aim, we consider the so-called dual functional. Let us denote by P_Λ^* the probability in Equation (3). Then, the dual functional, \mathcal{J} , which is a real valued function of the Lagrange multipliers, is defined by:

$$\mathcal{J}(\Lambda) := \mathcal{L}(P_\Lambda^*) \tag{4}$$

Observing that:

$$\Lambda(P_\Lambda^*) = \sum_{h,k} \lambda_{hk} c_{hk} - \sum_{\sigma} \sum_{h,k} \lambda_{hk} \sigma_h \sigma_k P_\Lambda^*(\sigma)$$

and:

$$\begin{aligned} \mathcal{S}(P_\Lambda^*) &= \log Z + \sum_{\sigma} \sum_{h,k} \lambda_{hk} \sigma_h \sigma_k P_\Lambda^*(\sigma) \\ &= \log \left[\sum_{\sigma} \exp \left\{ - \sum_{h,k} \lambda_{hk} \sigma_h \sigma_k \right\} \right] + \sum_{\sigma} \sum_{h,k} \lambda_{hk} \sigma_h \sigma_k P_\Lambda^*(\sigma) \end{aligned}$$

we obtain the convex function:

$$\mathcal{J}(\Lambda) = \sum_{h,k} \lambda_{hk} c_{hk} + \log \left[\sum_{\sigma} \exp \left\{ - \sum_{h,k} \lambda_{hk} \sigma_h \sigma_k \right\} \right] \tag{5}$$

If we denote by $\nabla \mathcal{J}$ the gradient of \mathcal{J} with respect to the variables, λ_{ij} , it is immediately seen that the following statements are equivalent:

- Λ is a critical point for \mathcal{J} , i.e., $\nabla \mathcal{J}(\Lambda) = 0$;
- P_Λ^* realizes the assigned correlations, i.e.,:

$$\sum_{\sigma} \sum_{h,k} \lambda_{hk} \sigma_h \sigma_k P_\Lambda^*(\sigma) = c_{hk}$$

for every h, k .

A critical point exists if \mathcal{J} is proper, which means:

$$\lim_{\|\Lambda\| \rightarrow \infty} \mathcal{J}(\Lambda) = +\infty$$

It is also clear that the following set of inequalities ensures the properness of \mathcal{J} :

$$\sum_{i,j} c_{ij} \lambda_{ij} > \min \left\{ \sum_{i,j} \lambda_{ij} \sigma_i \sigma_j : \sigma \in \Omega \right\}, \text{ for every } \Lambda$$

Let us denote by $M(\Lambda)$, the minimum given by: $\min \left\{ \sum_{i,j} \lambda_{ij} \sigma_i \sigma_j : \sigma \in \Omega \right\}$. We denote by Δ_n the set of matrices defined by:

$$\Delta_n = \left\{ C = (c_{ij}) : \sum_{i,j} c_{ij} \lambda_{ij} \geq M(\Lambda) \text{ for every } \Lambda \right\} \tag{6}$$

We can now state the main result of this section.

Theorem 2 *Let Δ_n be as defined in Equation (6). Then:*

$$\Delta_n = \text{Cov}_n$$

Proof. We first show that $\Delta_n \subseteq \text{Cov}_n$. Since Cov_n is closed, it is enough to show that $\overset{\circ}{\Delta}_n \subseteq \text{Cov}_n$, where $\overset{\circ}{\Delta}_n$ denotes the interior of Δ_n . We know that

$$\overset{\circ}{\Delta}_n = \left\{ C = (c_{ij}) : \sum_{i,j} c_{ij} \lambda_{ij} > M(\Lambda) \text{ for every } \Lambda \right\}$$

Thus, for $C \in \overset{\circ}{\Delta}_n$, the dual functional, $\mathcal{J}(\Lambda)$, is proper. This implies feasibility. Thus, there exists a probability, P , that realizes C as a correlation matrix of spin variables. Hence, $C \in \text{Cov}_n$.

Now, to show $\text{Cov}_n \subseteq \Delta_n$, let $C = (c_{ij}) \in \text{Cov}_n$. Then, for every Λ , we have:

$$\begin{aligned} \sum_{i,j} \lambda_{ij} c_{ij} &= E \left(\sum_{i,j} \lambda_{ij} \sigma_i \sigma_j \right) \\ &\geq \min_{\sigma} \sum_{i,j} \lambda_{ij} \sigma_i \sigma_j \\ &= M(\Lambda) \end{aligned}$$

This implies $C \in \Delta_n$. As pointed out by one reviewer, Theorem 4.1 can also be proven by contradiction using the hyperplane separation theorem, using the knowledge of the extremal points of the polytope. Our proof, however, does not rely on this knowledge and holds, with minimal modifications, for random variables taking values in general subsets of \mathbb{R} . The theorem above provides a (non-minimal) dual description of the polytope of spin-correlation matrices.

Its main consequence is that it guarantees that whenever C is in the interior of Cov_n , then it can be realized by some probability of the form Equation (3), for a Λ , which minimizes the dual functional \mathcal{J} . The search of a probability that realizes a given correlation matrix is therefore reduced to finding the minimum of a function that, as we will see shortly, is convex. Before giving some details on this point, we observe that various classes of inequalities are obtained from Equation (6) by a suitable choice of Λ .

- Positivity: let $x \in \mathbb{R}^n$ and set $\lambda_{ij} = x_i x_j$. Then, for every $\sigma \in \Omega_n$:

$$\sum_{ij} \lambda_{ij} \sigma_i \sigma_j = \sum_{ij} x_i x_j \sigma_i \sigma_j = \frac{1}{2} \left[\sum_i x_i \sigma_i \right]^2 \geq 0$$

So, $M(\Lambda) \geq 0$. Thus, for $C \in \Delta_n$

$$\sum_{ij} x_i x_j c_{ij} \geq 0$$

which implies positivity.

- Bell's inequalities: let $A \subset \{1, 2, \dots, n\}$ with $|A| = 3$ and $\varepsilon \in \{-1, 1\}^A$. We set:

$$\lambda_{ij} = \begin{cases} \varepsilon_i \varepsilon_j & \text{for } i, j \in A, i \neq j \\ 0 & \text{otherwise} \end{cases}$$

Then, for $A = \{r, s, t\}$:

$$\begin{aligned} \frac{1}{2} \sum_{ij} \lambda_{ij} \sigma_i \sigma_j &= \varepsilon_r \varepsilon_s \sigma_r \sigma_s + \varepsilon_r \varepsilon_t \sigma_r \sigma_t + \varepsilon_s \varepsilon_t \sigma_s \sigma_t \\ &= \eta_r \eta_s + \eta_r \eta_t + \eta_s \eta_t \geq -1 \end{aligned}$$

where $\eta_i = \varepsilon_i \sigma_i$. So, we have $M(\Lambda) \geq -1$. Thus, for $C \in \Delta_n$

$$\varepsilon_r \varepsilon_s c_{rs} + \varepsilon_r \varepsilon_t c_{rt} + \varepsilon_s \varepsilon_t c_{st} \geq -1$$

which are Bell's inequalities.

- Generalizations of Bell's inequalities: Let us consider $T \subset \{1, 2, \dots, n\}$, such that $|T|$ is odd. Then, let $\varepsilon \in \{-1, 1\}^T$. We set:

$$\lambda_{ij} = \begin{cases} \varepsilon_i \varepsilon_j & \text{for } i, j \in T, i \neq j \\ 0 & \text{otherwise} \end{cases}$$

we have:

$$\sum_{i,j} \lambda_{ij} \sigma_i \sigma_j = \left(\sum_{i \in T} \varepsilon_i \sigma_i \right)^2 - |T|$$

since, $|T|$ is odd, we have:

$$\min_{\sigma} \left(\sum_{i \in T} \varepsilon_i \sigma_i \right)^2 = 1$$

Thus, $M(\Lambda) = 1 - |T|$. As a result, we obtain the inequality

$$|T| - 1 + \sum_{i \neq j} \varepsilon_i \varepsilon_j c_{ij} \geq 0 \tag{7}$$

We call these the generalized Bell's inequalities. These, as kindly pointed out by one anonymous reviewer, are special instances of the hypermetric inequalities introduced by M. Deza in the 1960s, (Section 6.1 in [41]). They reduce to Bell's inequalities for $|T| = 3$. We have seen in the previous section that these inequalities, for $|T| = 3$ and $|T| = 5$, give the facets of the polytope of spin-correlation matrices for $n = 5$.

Many other variants of the Bell’s inequalities could be obtained with other choices of the λ_{ij} . For instance, we can generalize to all even dimensions the inequalities of type (three) for the case $n = 6$. Let $n \geq 6$ be even, and consider $T \subset \{1, 2, \dots, n\}$, such that $|T| = n - 1$. Then, choose:

$$\lambda_{ij} = \begin{cases} \varepsilon_i \varepsilon_j & \text{for } i, j \in T, i \neq j \\ 2\varepsilon_i \varepsilon_j & \text{for } i \in T, j \in T^c \text{ or } i \in T^c, j \in T \\ 0 & \text{otherwise} \end{cases}$$

We obtain:

$$\sum_{i,j} \lambda_{ij} \sigma_i \sigma_j = \left(\sum_{i \in T} \varepsilon_i \sigma_i \right)^2 - |T| + 4\varepsilon_{j_T} \left(\sum_{i \in T} \varepsilon_i \sigma_i \right)$$

where j_T is the only element of T^c . It is easy to check that the expression:

$$\left(\sum_{i \in T} \varepsilon_i \sigma_i \right)^2 + 4\varepsilon_{j_T} \left(\sum_{i \in T} \varepsilon_i \sigma_i \right)$$

as a function of $\sigma \in \Omega_n$, attains its minimum at $\left(\sum_{i \in T} \varepsilon_i \sigma_i \right) = k\varepsilon_{j_T}$ with k equal to three or five, and the minimum is -3 , which gives $M(\Lambda) = -(|T| + 3) = -(n + 2)$, and the family of inequalities:

$$n + 2 + \sum_{i \neq j; i, j \in T} \varepsilon_i \varepsilon_j c_{ij} + 4 \sum_{i \in T} \varepsilon_i \varepsilon_{j_T} c_{ij_T} \geq 0$$

These, for $n = 6$, reduce to the inequality of type (three).

Remark 1 *It is important to note that nowhere in the process of obtaining the maximum entropy measure have we assumed that we are given all the c'_{ij} s. Suppose we are only given a partial matrix. Then, Δ_n can be interpreted as the set of conditions under which the given partial matrix can be extended to a spin correlation matrix. Once we have feasibility, we know that the maximum entropy measure, P^* , exists and can be used to complete the given matrix to a spin correlation matrix.*

4.2. Finding the Minimum of the Dual Functional

We have observed that the maximum entropy method allows us to reduce the problem of realizing a given spin correlation matrix to finding the minimum of the function, \mathcal{J} , defined in Equation (5). In this section, we show that this minimum can be obtained by an explicit gradient descent algorithm. Note first that \mathcal{J} has some obvious symmetry properties: $\mathcal{J}(\Lambda) = \mathcal{J}(\Lambda')$ if $\lambda_{ij} = \lambda'_{ij}$ for all $i \neq j$, and \mathcal{J} is indifferent to symmetrization:

$$\mathcal{J}(\Lambda) = \mathcal{J} \left(\frac{\Lambda + \Lambda^T}{2} \right)$$

where A^T is the transposition of the matrix, A . It is, therefore, enough to deal with the minimization problem within the set of symmetric matrices with zero diagonal elements. These matrices can be identified with elements of \mathbb{R}^I , where:

$$I := \{(i, j) : 1 \leq i < j \leq n\}$$

In what follows, we use the usual vector notation for elements of \mathbb{R}^I : for $v, w \in \mathbb{R}^I$, v^T denotes its transposition, $v^T w$ is the scalar product in \mathbb{R}^I and vw^T is an element of $\mathbb{R}^{I \times I}$.

Proposition 1 Consider the discrete time dynamical system in \mathbb{R}^I , defined by:

$$\lambda(t + 1) = \lambda(t) - \frac{1}{K} \nabla \mathcal{J}(\lambda(t)) \tag{8}$$

For every $K > \frac{n(n-1)}{2}$, this system has a unique fixed point, λ^* , which is a global attractor, and it is the unique minimum of \mathcal{J} .

Proof. Let $G : \{-1, 1\}^n \rightarrow \mathbb{R}^I$ be defined by:

$$G_{ij}(\sigma) := \sigma_i \sigma_j$$

for $(i, j) \in I$. Moreover, $C \in Cov_n$ are also obviously identified with elements of \mathbb{R}^I . In particular, in what follows, C^T denotes the transposition of C as a vector in \mathbb{R}^I , rather than a matrix in $M_n(\mathbb{R})$. With these notations, \mathcal{J} can be rewritten as:

$$\mathcal{J}(\lambda) = C^T \lambda + \log \left(\sum_{\sigma} e^{G^T(\sigma)\lambda} \right)$$

By elementary computations, we can compute the gradient, $\nabla \mathcal{J}$, and the Hessian, $\nabla^2 \mathcal{J}$:

$$\begin{aligned} \nabla \mathcal{J}(\lambda) &= C^T - \frac{\sum_{\sigma} G^T e^{G^T(\sigma)\lambda}}{\sum_{\sigma} e^{G^T(\sigma)\lambda}} \\ \nabla^2 \mathcal{J}(\lambda) &= \frac{\sum_{\sigma} G G^T e^{G^T(\sigma)\lambda}}{\sum_{\sigma} e^{G^T(\sigma)\lambda}} - \frac{\sum_{\sigma} G e^{G^T(\sigma)\lambda}}{\sum_{\sigma} e^{G^T(\sigma)\lambda}} \frac{\sum_{\sigma} G^T e^{G^T(\sigma)\lambda}}{\sum_{\sigma} e^{G^T(\sigma)\lambda}} \end{aligned}$$

It follows, in particular, that $\nabla^2 \mathcal{J}(\lambda)$ is the covariance matrix of the vector, G , with respect to the probability:

$$P_{\lambda}^*(\sigma) = \frac{e^{G^T(\sigma)\lambda}}{\sum_{\sigma'} e^{G^T(\sigma')\lambda}} \tag{9}$$

and it is, therefore, nonnegative. Thus, \mathcal{J} is convex. In the next steps, we establish more detailed properties of \mathcal{J} , including its strict convexity.

Step 1: the elements of G are linearly independent functions. Suppose α_0, α_{ij} for $(i, j) \in I$, such that:

$$\alpha_0 + \sum_{1 \leq i < j \leq n} \alpha_{ij} \sigma_i \sigma_j = 0 \tag{10}$$

for every $\sigma \in \{-1, 1\}^n$. We show that:

$$\alpha_0 = \alpha_{ij} = 0 \tag{11}$$

for every $(i, j) \in I$. We proceed by induction on n . There is nothing to prove for $n = 1$. We can write, assuming Equation (10):

$$\sigma_1 \sum_{j=2}^n \alpha_{1j} \sigma_j + \left(\alpha_0 + \sum_{2 \leq i < j \leq n} \alpha_{ij} \sigma_i \sigma_j \right) = 0 \tag{12}$$

Since the second summand in Equation (12) does not contain σ_1 , this implies:

$$\sum_{j=2}^n \alpha_{1j} \sigma_j \equiv 0 \tag{13}$$

and:

$$\alpha_0 \sum_{2 \leq i < j \leq n} \alpha_{ij} \sigma_i \sigma_j \equiv 0 \tag{14}$$

Identity in Equation (13) implies $\alpha_{1j} = 0$ for $j = 2, \dots, n$, as can be shown, for instance, again by induction on n . Identity in Equation (14) implies $\alpha_0 = \alpha_{ij} = 0$ for $2 \leq i < j \leq n$ by the inductive assumption.

Step 2: for every $\lambda \in \mathbb{R}^I$, $\nabla^2 \mathcal{J}(\lambda)$ is strictly positive definite. Denote by $\mathbb{V}(G)$ the covariance matrix of G with respect to the probability in Equation (9). We have:

$$\nabla^2 \mathcal{J}(\lambda) = \mathbb{V}(GG^T)$$

Thus, for $b \in \mathbb{R}^I$:

$$b^T \nabla^2 \mathcal{J}(\lambda) b = 0 \Rightarrow \mathbb{V}(b^T G) = 0$$

that, since P_λ^* is fully supported, gives:

$$b^T G(\sigma) = \text{constant}$$

for every $\sigma \in \{-1, 1\}^n$. By Step 1, this implies that $b = 0$, which proves the claim.

Step 3: $\lambda \in \mathbb{R}^I$, the largest eigenvalue of $\nabla^2 \mathcal{J}(\lambda)$ is less than or equal to $\frac{n(n-1)}{2}$. Let δ denote this largest eigenvalue and $v \in \mathbb{R}^I$, a corresponding eigenvector with $v^T v = 1$. We have:

$$\begin{aligned} \delta &= v^T \nabla^2 \mathcal{J}(\lambda) v = \mathbb{V}(v^T G) \\ &\leq \sup_{\sigma} [v^T G(\sigma)]^2 \leq \sup_{\sigma} G^T(\sigma) G(\sigma) = |I| = \frac{n(n-1)}{2} \end{aligned}$$

Step 4: For $K > \frac{n(n-1)}{2}$, the map, $\lambda \mapsto \lambda - \frac{1}{K} \nabla \mathcal{J}(\lambda)$, is a strict contraction, and therefore, it has a unique fixed point. Let:

$$\phi(\lambda) := \lambda - \frac{1}{K} \nabla \mathcal{J}(\lambda)$$

We have:

$$\phi(\lambda) - \phi(\mu) = \lambda - \mu - \frac{1}{K} \nabla^2 \mathcal{J}(\xi)(\lambda - \mu)$$

for some ξ in the segment joining λ and μ . Thus, setting $\|v\|^2 = v^T v$:

$$\|\phi(\lambda) - \phi(\mu)\|^2 = \left\| \left[I - \frac{1}{K} \nabla^2 \mathcal{J}(\xi) \right] (\lambda - \mu) \right\|^2$$

The conclusion now follows from the fact that, by Steps 2 and 3, the symmetric matrix, $I - \frac{1}{K} \nabla^2 \mathcal{J}(\xi)$, has all eigenvalues in $(0, 1)$.

Step 5: Conclusion. By Step 4, the system, Equation (8), has a unique fixed point, λ , for which, necessarily, $\nabla \mathcal{J}(\lambda) = 0$.

It should be observed that the computation of $\mathcal{J}(\lambda)$ and of its gradient involves computing a sum over $\sigma \in \{-1, 1\}^n$. This may be hard or even practically unfeasible for large n ; this difficulty may be made less severe by the use of Monte Carlo methods. This and other computational aspects of this algorithm will be discussed in a forthcoming paper.

Acknowledgments

We are grateful to an anonymous reviewer for pointing out to us some literature in discrete geometry that is relevant for the present paper. The work of M. Pavon was partially supported by the QuantumFuture research grant of the University of Padova and by an Alexander von Humboldt Foundation fellowship at the Institut für Angewandte Mathematik, Universität Heidelberg, Germany.

Conflict of Interest

The authors declare no conflict of interest.

References

1. Stieltjes, T.J. Recherches sur les fractions continues (in French). *Annales de la Facult des Sciences de Toulouse* **1894**, *8*, J1–J122.
2. Akhiezer, N.I. *The Classical Moment Problem and Some Related Questions in Analysis (Translated from Russian by N. Kemmer)*; Hafner Publishing Co.: New York, NY, USA, 1965.
3. Krein, M.G.; Nudelman, A.A. *The Markov Moment Problem and Extremal Problems*; American Mathematical Society: Providence, RI, USA, 1977.
4. Grenander, U.; Szegö, G. *Toeplitz Forms and Their Applications*; University of California Press: Berkeley, CA, USA, 1958.
5. Sarason, D. Generalized interpolation in H^∞ . *Trans. Am. Math. Soc.* **1967**, *127*, 179–203.
6. Arov, D.Z.; Dym, H. On three Krein extension problems and some generalizations. *Int. Equat. Oper. Theory* **1998**, *31*, 1–91.
7. Blomqvist, A.; Lindquist, A.; Nagamune, R. Matrix-valued Nevanlinna-Pick interpolation with complexity constraint: An optimization approach. *IEEE Trans. Autom. Control* **2003**, *48*, 2172–2190.
8. Georgiou, T. The interpolation problem with a degree constraint. *IEEE Trans. Autom. Control* **1999**, *44*, 631–635.
9. Georgiou, T. Realization of power spectra from partial covariance sequences. *IEEE Trans. Acoust. Speech Signal Process.* **1987**, *35*, 438–449.
10. Byrnes, C.; Lindquist, A.; Gusev, S.; Matveev, A.S. A complete parameterization of all positive rational extensions of a covariance sequence. *IEEE Trans. Autom. Control* **1995**, *40*, 1841–1857.
11. Byrnes, C.; Lindquist, A. On the partial stochastic realization problem. *IEEE Trans. Autom. Control* **1997**, *42*, 1049–1070.
12. Byrnes, C.I.; Gusev, S.; Lindquist, A. A convex optimization approach to the rational covariance extension problem. *SIAM J. Control Optim.* **1999**, *37*, 211–229.

13. Byrnes, C.I.; Georgiou, T.; Lindquist, A. A generalized entropy criterion for Nevanlinna-Pick interpolation with degree constraint: A convex optimization approach to certain problems in systems and control. *IEEE Trans. Autom. Control* **2001**, *46*, 822–839.
14. Georgiou, T. Spectral analysis based on the state covariance: The maximum entropy spectrum and linear fractional parameterization. *IEEE Trans. Autom. Control* **2002**, *47*, 1811–1823.
15. Georgiou, T.; Lindquist, A. Kullback-Leibler approximation of spectral density functions. *IEEE Trans. Inf. Theory* **2003**, *49*, 2910–2917.
16. Georgiou, T. The structure of state covariances and its relation to the power spectrum of the input. *IEEE Trans. Autom. Control* **2002**, *47*, 1056–1066.
17. Georgiou, T. Solution of the general moment problem via a one-parameter imbedding. *IEEE Trans. Autom. Control* **2005**, *50*, 811–826.
18. Byrnes, C.I.; Linquist, A. Important moments in systems and control. *SIAM J. Control Opt.* **2008**, *47*, 2458–2469.
19. Ferrante, A.; Pavon, M.; Ramponi, F. Hellinger vs. Kullback-Leibler multivariable spectrum approximation. *IEEE Trans. Autom. Control* **2008**, *53*, 954–967.
20. Ramponi, F.; Ferrante, A.; Pavon, M. A globally convergent matricial algorithm for multivariate spectral estimation. *IEEE Trans. Autom. Control* **2009**, *54*, 2376–2388.
21. Ferrante, A.; Pavon, M.; Zorzi, M. A maximum entropy enhancement for a family of high-resolution spectral estimators. *IEEE Trans. Autom. Control* **2012**, *57*, 318–329.
22. Ferrante, A.; Masiero, C.; Pavon, M. Time and spectral domain relative entropy: A new approach to multivariate spectral estimation. *IEEE Trans. Autom. Control* **2012**, *57*, 2561–2575.
23. Byrnes, C.I.; Enqvist, P.; Linquist, A. Identifiability and well-posedness of shaping-filter parameterizations: A global analysis approach. *SIAM J. Control Optim.* **2002**, *41*, 23–59.
24. Georgiou, T.; Lindquist, A. A convex optimization approach to ARMA modeling. *IEEE Trans. Autom. Control* **2008**, *AC-53*, 1108–1119.
25. Enqvist, P.; Karlsson, J. Minimal Itakura-Saito Distance and Covariance Interpolation. In Proceedings of the 47th IEEE Conference on Decision and Control (CDC 2008), Cancun, Mexico, 9–11 December 2008; pp. 137–142.
26. Byrnes, C.I.; Georgiou, T.; Lindquist, A.; Megretski, A. Generalized interpolation in H-infinity with a complexity constraint. *Trans. Am. Math. Soc.* **2006**, *358*, 965–987.
27. Georgiou, T.; Lindquist, A. Remarks on control design with degree constraint. *IEEE Trans. Autom. Control* **2006**, *AC-51*, 1150–1156.
28. Nasiri Amini, A.; Ebbini, E.; Georgiou, T. Noninvasive estimation of tissue temperature via high-resolution spectral analysis techniques. *IEEE Trans. Biomed. Eng.* **2005**, *52*, 221–228.
29. Dempster, A.P. Covariance selection. *Biometric* **1972**, *28*, 157–175.
30. Burg, J.; Luenberger, D.; Wenger, D. Estimation of structured covariance matrices. *Proc. IEEE* **1982**, *70*, 963–974.
31. Dembo, A.; Mallows, C.; Shepp, L.A. Embedding nonnegative definite Toeplitz matrices in nonnegative definite circulant matrices, with application to covariance estimation. *IEEE Trans. Inf. Theory* **1989**, *35*, 1206–1212.

32. Barrett, W.; Johnson, C.; Loewy, R. The real positive definite completion problem: Cycle completability. *Mem. Am. Math. Soc.* **1996**, *122*, 1–68.
33. Laurent, M. The real positive semidefinite completion problem for series-parallel graphs. *Linear Algebra Appl.* **1997**, *252*, 347–366.
34. Glunt, W.; Hayden, T.; Johnson, C.; Tarazaga, P. Positive definite completions and determinant maximization. *Linear Algebra Appl.* **1999**, *288*, 1–10.
35. Laurent, M. Polynomial instances of the positive semidefinite and Euclidean distance matrix completion problems. *SIAM J. Matrix Anal. Appl.* **2001**, *22*, 874–894.
36. Ferrante, A.; Pavon, M. Matrix completion à la Dempster by the principle of parsimony. *IEEE Trans. Inf. Theory* **2011**, *57*, 3925–3931.
37. Carli, F.; Ferrante, A.; Pavon, M.; Picci, G. A maximum entropy solution of the covariance extension problem for reciprocal processes. *IEEE Trans. Autom. Control* **2011**, *56*, 1999–2012.
38. Parthasarathy, K.R. On extremal correlations. *J. Stat. Plan. Inference* **2002**, *102*, 282–285.
39. Gupta, J.C. Characterisation of correlation matrices of spin variables. *Sankhya Indian J. Stat.* **1999**, *61*, 282–285.
40. Pitowsky, I. Correlation polytopes: Their geometry and complexity. *Math. Program.* **1991**, *50*, 395–414.
41. Deza, M.M.; Laurent, M. *Geometry of Cuts and Metrics*; Springer: Berlin, Germany, 1997.
42. Balasubramanian, K.; Gupta, J.C.; Parthasarathy, K.R. Remarks on Bell's inequalities for spin correlations. *Sankhya Indian J. Stat.* **1998**, *60*, 29–35.
43. Werner, R.F.; Wolf, M.M. Bell inequalities and entanglement. *Quant. Inf. Comput.* **2001**, *1*, 1–25.
44. Kuna, T.; Lebowitz, J.L.; Speer, E.R. Necessary and sufficient conditions for realizability of point processes. *Ann. Appl. Probab.* **2011**, *11*, 1253–1281.
45. Dai Pra, P.; Pavon, M.; Sahasrabudhe, N. A note on the geometric interpretation of Bell's inequalities. *Lett. Math. Phys.* **2013**, Preprint:arXiv:1301.4823v3. To appear.
46. McRae, W.B.; Davidson, E.R. Linear inequalities for density matrices II. *J. Math. Physics* **1972**, *13*, 1527–1538.
47. Bremner, D.; Fukuda, K.; Marzetta, A. Primal-dual methods for vertex and facet enumeration. *Discret. Comput. Geom.* **1998**, *20*, 333–357.
48. Fukuda, K.; Prodon, A. Double Description Method Revisited. In *Combinatorics and Computer Science (Brest, 1995)*; Springer: Berlin, Germany, 1996; Volume 1120, *Lecture Notes in Computer Science*, pp. 91–111.
49. Jaynes, E.T. Information theory and statistical mechanics. *Phys. Rev. Ser. II* **1957**, *106*, 620–630.
50. Jaynes, E.T. Information theory and statistical mechanics II. *Phys. Rev. Ser. II* **1957**, *108*, 171–190.
51. Jaynes, E.T. On the rationale of maximum-entropy methods. *Proc. IEEE* **1982**, *70*, 939–952.
52. Csiszár, I. Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems. *Ann. Stat.* **1991**, *19*, 2032–2066.