Matrix Completion à la Dempster by the Principle of Parsimony

Augusto Ferrante and Michele Pavon

Abstract—Dempster’s covariance selection method is extended first to general nonsingular matrices and then to full rank rectangular matrices. Dempster observed that his completion solved a maximum entropy problem. We show that our generalized completions are also solutions of a suitable entropy-like variational problem.

Index Terms—Covariance selection, matrix completion, maximum entropy problem, parsimony principle, variational problem.

I. DEMPSTER’S COVARIANCE SELECTION

In the seminal paper [16], Dempster introduced a general strategy for completing a partially specified covariance matrix. Consider a zero-mean, multivariate Gaussian distribution with density

\[ p(x) = (2\pi)^{-n/2} |\Sigma|^{-1/2} \exp \left\{ -\frac{1}{2} x^\top \Sigma^{-1} x \right\}, \quad x \in \mathbb{R}^n. \]

Suppose that the elements \( \sigma_{ij} \), \( 1 \leq i \leq j \leq n \), \((i, j) \in \mathcal{I}\) have been specified. How should \( \Sigma \) be completed? Dempster resorts to a form of the Principle of Parsimony in parametric model fitting: As the elements \( \sigma_{ij} \) of \( \Sigma^{-1} \) appear as natural parameters of the model, one should set \( \sigma_{ij} = 0 \) to zero for \( 1 \leq i \leq j \leq n \), \((i, j) \notin \mathcal{I}\). Notice that \( \sigma_{ij} = 0 \) has the probabilistic interpretation that the \( i \)th and \( j \)th components of the Gaussian random vector are conditionally independent given the other components. This choice, which we name henceforth Dempster’s Completion, may at first look less natural than setting the unspecified elements \( \Sigma \) to zero. It has nevertheless considerable advantages compared to the latter, cf. [16, p.161]. In particular, Dempster established the following far reaching result.

Theorem 1: Assume that a symmetric, positive-definite completion of \( \Sigma \) exists. Then there exists a unique Dempster’s Completion \( \Sigma^o \). This completion maximizes the (differential) entropy

\[ H(p) = -\int_{\mathbb{R}^n} \log(p(x)) p(x) dx \]

\[ = \frac{1}{2} \log(\det \Sigma) + \frac{1}{2} n (1 + \log(2\pi)) \]  

Manuscript received June 28, 2010; accepted November 24, 2010. Date of current version May 25, 2011. This work was supported in part by the Italian Ministry for Education and Research (MIUR) under PRIN Grant “Identification and Robust Control of Industrial Systems,” in part by the CPDA080209/08, in part by the QIFENE research grants of the University of Padova, and in part by the Department of Information Engineering research project “QUINTET.” A. Ferrante is with the Dipartimento di Ingegneria dell’Informazione, Università di Padova, 35131 Padova, Italy (e-mail: augusto@dei.unipd.it).

M. Pavon is with the Dipartimento di Matematica Pura ed Applicata, Università di Padova, 35131 Padova, Italy (e-mail: pavon@math.unipd.it).

Communicated by J. Romberg, Associate Editor for Signal Processing.

Digital Object Identifier 10.1109/TIT.2011.2143970

among zero-mean Gaussian distributions having the prescribed elements \( \{\sigma_{ij} \mid 1 \leq i \leq j \leq n, (i, j) \in \mathcal{I}\} \).

Thus, Dempster’s Completion \( \Sigma^o \) solves a maximum entropy problem, i.e., maximizes entropy under linear constraints. Dempster’s paper has generated a whole stream of research, see e.g. [6], [15], [23], [35] and references therein. In the meantime, matrix completion has become an important area of research with several new applications, where the completed matrix must have certain prescribed properties: For instance, it should be positive definite, it should be circulant, it should have a Toeplitz structure, it should have a prescribed low rank, etc. Motivation originates from problems in texture images modeling, recommender systems and networked sensors [4], [10]–[13], [22], [26], [28], [29], [31].

In this paper, we consider a totally unstructured version of Dempster’s problem. A (possibly non square) matrix \( M \) is partially specified and we seek to complete it according to a generalized principle of parsimony. We are motivated by the following applications:

• Covariance selection: There exist a number of interesting variants:
  — Dempster’s original covariance completion problem;
  — if the covariance matrix originates from a stationary time series, then the completion should feature a Töplitz structure [6];
  — if it originates from a (vector) process stationary on the discrete circle, then the completion should have a (block) Töplitz and circulant structure, [12], [13].

• Newton’s step: Suppose we are interested in finding the critical points of the twice differentiable function \( f : \mathbb{R}^n \rightarrow \mathbb{R} \). Consider the basic iterative scheme of Newton’s method:

\[ x_{n+1} = x_n - [Hf(x_n)]^{-1} \nabla f(x_n), \quad n \geq 0 \]  

where \( Hf(x_n) \) is the Hessian of \( f \) at \( x_n \) and \( \nabla f(x_n) \) is the gradient of \( f \) at \( x_n \). As is well-known, the most expensive part of computing the Newton step \( [Hf(x_n)]^{-1} \nabla f(x_n) \) is finding, storing and inverting the Hessian. For problems with special structure, sometimes one spares computing a portion of the Hessian (at least in some steps). It those cases, the Hessian needs to be suitably completed to a symmetric matrix. In the case of saddle-point search for a Lagrangian, the Hessian is indefinite.

• Linear systems of equations: Consider the linear system

\[ MX = B \]

where the matrix \( B \) is given. Such systems occur in almost every branch of science and engineering. Suppose
that some of the elements of the square matrix $M$ are missing. For instance, in an open Leontief input-output model [30], the goods production of some industries might not have been decided yet. Similarly, in designing an electrical network obeying Ohm’s and Kirchhoff’s laws [36], the resistance of some resistors might not have been decided/determined. In the absence of any rank information, as a generic completion $M_0$ is invertible, we can associate to such a completion the solution

$$X_{M_0} = M_0^{-1} B.$$  \hfill (4)

In all of the above examples, a partially known matrix needs to be completed, but it is the inverse of the matrix that is actually needed in the probability density, in the Newton step and in solving the linear system. Motivated by this observation and by Dempster’s classical results, we identify as desirable, according to a generalized principle of parsimony, completions that feature a maximum number of zeros in the inverse. Notice that such completions in (2) and (4) lead to a reduced computational burden. In this paper, we show that a family of such desirable completions are generalized Dempster’s completions that can be characterized as critical points of a suitable variational problem.

It may, at first, look hopeless to obtain an entropy-like variational characterization of Dempster’s-like completions without positivity. Nevertheless, we prove in Lemmata 1 and 2 below that the constrained extremization of the determinant only involves the positive part of the matrix $(MM^T)^{1/2}$. More precisely, only the singular values of $M$ come into play. Hence, such a variational characterization is possible and may be established even in the rectangular case.

The paper is outlined as follows. We discuss first the square case to facilitate the comparison with Dempster’s classical results, see Sections III and IV below. In Section V, we discuss two examples to illustrate the properties that our solutions may or may not enjoy. In Section VI, we generalize our results to rectangular, full rank matrices. The paper concludes with a discussion section comparing our approach to other matrix completion techniques and to other moment problems.

II. THE GENERAL COMPLETION PROBLEM

Let $\mathcal{I} \subset \{1, 2, \ldots, n\} \times \{1, 2, \ldots, n\}$ and $\bar{\mathcal{I}}$ be the complementary subset. To each $(i, j) \in \mathcal{I}$ we associate the unknown $x_{ij}$. Let $x$ be the vector, say $k$-dimensional, obtained by stacking the $x_{ij}$ one on top of the other. We define as partial matrix a parameteric family of matrices $M(x)$ whose entries $[M(x)]_{ik} = m_{ik}$ are specified for $(i, j) \in \mathcal{I}$, while $[M(x)]_{ik} = x_{ij}$ for $(i, j) \in \bar{\mathcal{I}}$. Here, both $m_{ik}$ and $x_{ij}$ take real values. A completion of the partial matrix $M(x)$ is a matrix $M(x_0)$ where $x_0 \in \mathbb{R}^k$. Notice that completions always exist as we are not requiring $M(x_0)$ to possess any further property. If $M(x)$ is a partial matrix and $\mathcal{I}$ is the corresponding set of indices of the unspecified elements, we denote by $\mathcal{I}^\perp$ the set of indices $\mathcal{I}^\perp := \{(j, i) : (i, j) \in \mathcal{I}\}$.

Let $M(x)$ be a square partial matrix of size $n$ and let $\mathcal{I}$ be the corresponding set of indices of the unspecified entries. Consider the following matrix completion problems:

**Problem 1:** Find the nonsingular completions $M(x_0)$ such that $[M(x_0)^{-1}]_{ij} = 0$ for all $(i, j) \in \mathcal{I}^\perp$.

**Remark:** Since in $M(x)$ we have $|\mathcal{I}|$ degrees of freedom (unknowns), we may generically expect that the maximum number of entries that can be annihilated in $M(x)^{-1}$ is precisely $|\mathcal{I}|$. Indeed, $m$ zeros in $M(x)^{-1}$ is equivalent to finding a solution to a system of $m$ polynomial equations in $|\mathcal{I}|$ unknown on the complement of the set where the determinant vanishes (in Section V, we provide a non generic example where there is a completion with more than $|\mathcal{I}|$ zeros in the inverse).

**Problem 2:** Find the nonsingular completions $M(x_0)$ that extremize $\det[M(x)]$.

**Remark:** Notice that when a covariance matrix $\Sigma = \Sigma^T > 0$ is sought, Problem 2 reduces to the maximum entropy problem solved by Dempster’s Completion. This follows from the fact that the entropy, in the Gaussian case, differs from $(1/2) \log \det \Sigma$ by a constant, the monotonicity of the logarithm and strict concavity of the entropy. Thus, Problem 2 appears as a legitimate generalization of Dempster’s classical completion method.

Even without positivity, maximizing the determinant of a matrix may represent a natural goal. Consider, for instance, the Mueller matrix characterizing the polarization properties in optical systems [21]. In earth-space communications, the Mueller matrix of the optical space channel has to be estimated from noisy data [1]. Sometimes, some of the elements of the matrix cannot be estimated and are therefore missing. Maximizing the determinant of this matrix has then the significance of minimizing the noise effects [37].

The main result of this paper consists in showing that Problems 1 and 2 have the same set of solutions. Thus, the above example from optics provides further motivation for our study.

**Theorem 2:** $M(x_0)$ solves Problem 1 if and only if it solves Problem 2.

**Remark:** It is apparent that solutions to Problems 1 and 2 may not exist, but when they do there may be many. For instance, consider

$$M(x) = \begin{pmatrix} 1 & x_{12} \\ 0 & x_{22} \end{pmatrix}.$$  \hfill (5)

Then, Problems 1 and 2 are not solvable. Actually, whenever all the unknowns $x_{ij}$ are in the same row or in the same column, $\det(M(x))$ is linear in $x_{ij}$ and hence it does not have critical points. Two examples where there exist multiple solutions are provided in Section V.

III. SOME PRELIMINARY RESULTS

We collect below some lemmata that are needed to prove Theorem 2.

**Lemma 1:** Problem 2 is equivalent to the following: Find a nonsingular completion $M(x_0)$ that extremizes $J(M(x)) := \log |\det[M(x)]|$.
**Proof**: Compute the gradient

\[
\frac{\partial}{\partial x_t} \log |\det [M(x)]| = \frac{1}{\det [M(x)]} \frac{\partial |\det [M(x)]|}{\partial x_t}
\]

\[
= \frac{1}{\det [M(x)]} \frac{\partial}{\partial x_t} |\det [M(x)]|.
\]

(5)

Now the statement follows by observing that we are restricting attention to nonsingular completions. □

Denote by \(D[J(M); \delta M] \) the directional derivative of \(J \) in direction \(\delta M \in \mathbb{R}^{m \times n} \):

\[
D[J(M); \delta M] = \lim_{\varepsilon \to 0} \frac{\log |\det [M + \varepsilon \delta M]| - \log |\det [M]|}{\varepsilon}
\]

We have the following result.

**Lemma 2**: Let \(J(M(x)) = \log |\det [M(x)]| \) as in the previous lemma. If \(M \) is nonsingular then, for any \(\delta M \in \mathbb{R}^{m \times n} \),

\[
D[J(M); \delta M] = \text{tr} [M^{-1} \delta M].
\]

(7)

**Proof**: We have

\[
D[J(M); \delta M] = \lim_{\varepsilon \to 0} \frac{\log |\det [M + \varepsilon \delta M]| - \log |\det [M]|}{\varepsilon}
\]

\[
= \lim_{\varepsilon \to 0} \frac{\log |\det [I + \varepsilon M^{-1} \delta M]|}{\varepsilon}
\]

\[
= \lim_{\varepsilon \to 0} \frac{\log |\det [I + \varepsilon (\lambda_i \delta \lambda_i)]|}{\varepsilon}
\]

\[
= \lim_{\varepsilon \to 0} \frac{\sum_i \log |1 + \varepsilon \lambda_i|}{\varepsilon}
\]

\[
= \sum_i \lambda_i
\]

(12)

where the \(\lambda_i \)'s are the eigenvalues (counted with multiplicity) of \(M^{-1} \delta M \), so that their sum is indeed \(\text{tr} [M^{-1} \delta M] \). □

A more abstract argument permits to generalize this result to nonsquare matrices, see Lemma 4.

**Lemma 3**: Consider the space \(\mathbb{R}^{m \times n} \) endowed with the inner product \(\langle M_1, M_2 \rangle := \text{tr} [M_1^T M_2] \). Let \(\mathcal{M} \) be the subspace of \(\mathbb{R}^{m \times n} \) consisting of the matrices whose entries in position \((i,j) \) are zero. Let \(M \in \mathcal{M} \). Then, \([M]_{i,j} = 0 \) for all \((i,j) \in \mathcal{I} \).

**Proof**: Denote by \(e_i \) the \(i\)th canonical vector in \(\mathbb{R}^n \). Clearly, for any \((i,j) \in \mathcal{I}, e_i e_j^T \in \mathcal{M} \). Thus, if \(M \in \mathcal{M} \), for any \((i,j) \in \mathcal{I}, \) we have \(0 = \text{tr} [(e_i e_j^T)^T M] = \text{tr} [e_j e_i^T M] = e_i^T M e_j = [M]_{i,j} \). □

IV. PROOF OF THEOREM 2

We are now ready to prove our main result.

**Proof of Theorem 2**: In view of Lemma 1, Problem 2 is equivalent to the following variational problem:

\[
\text{extremize } \{ J(M) : e_i^T M e_j = m_{ij}, (i,j) \in \mathcal{I} \},
\]

(13)

where, as before, \(J(M(x)) = \log |\det [M(x)]| \). The corresponding Lagrangian is

\[
L(M) = J(M) + \sum_{(i,j) \in \mathcal{I}} \lambda_{ij} (m_{ij} - e_i^T M e_j)
\]

(14)

whose unconstrained extremization is obtained by annihilating the directional derivative of \(L \) in any direction \(\delta M \in \mathbb{R}^{n \times n} \). In view of Lemma 2, this yields:

\[
\text{tr} \left[ (M^{-1} - \sum_{(i,j) \in \mathcal{I}} \lambda_{ij} e_i e_j^T) \delta M \right] = 0, \quad \forall \delta M \in \mathbb{R}^{n \times n}
\]

(15)

or, equivalently,

\[
M^{-1} = \sum_{(i,j) \in \mathcal{I}} \lambda_{ij} e_i e_j^T.
\]

(16)

It follows, in particular, that the inverse of any nonsingular critical point \(M \) of the Lagrangian (14) has zeros in positions \((i,j) \in \mathcal{I} \). Moreover, if we can find \(\lambda_{ij} \) such that the matrix \(\sum_{(i,j) \in \mathcal{I}} \lambda_{ij} e_i e_j^T \) is nonsingular and

\[
M^0 := \left( \sum_{(i,j) \in \mathcal{I}} \lambda_{ij} e_i e_j^T \right)^{-1}
\]

(17)

satisfies

\[
e_i^T M^0 e_j = m_{ij}, \quad (i,j) \in \mathcal{I}
\]

(18)

then \(M^0 \) is indeed a solution of Problem 2.

Assume now that \(M(x_0) \) solves Problem 1. Then, \([M(x_0)]^{-1} \) has the form (16). Moreover, \(M(x_0) \) satisfies (18). Hence, it solves Problem 2.

Conversely, let \(M(x_0) \) solve Problem 2. This is equivalent to

\[
D[J(M), \delta M]_{M=M(x_0)} = \text{tr} [M(x_0)^{-1} \delta M] = (M(x_0)^{-1})^T, \delta M
\]

\[
= 0, \quad \forall \delta M \in \mathcal{M}
\]

(19)

where \(\mathcal{M} \) is the subspace of \(\mathbb{R}^{n \times n} \) (defined in Lemma 3) consisting of matrices whose entries in position \((i,j) \in \mathcal{I} \) are zero. Thus, \(M(x_0)^{-1} \in \mathcal{M} \). By Lemma 3, \([M(x_0)^{-1}]_{i,j} = 0 \) for all \((i,j) \in \mathcal{I} \), namely \(M(x_0) \) solves Problem 1. □

**Remark**: The last calculation in the above proof shows why, in Problem 1, \(\mathcal{I} \) rather than \(\mathcal{F} \) appears. Indeed, this follows from the orthogonality condition

\[
\langle M(x_0)^{-1}, \delta M \rangle = \text{tr} [M(x_0)^{-1} \delta M] = 0, \quad \forall \delta M \in \mathcal{M},
\]

V. TWO ILLUSTRATIVE EXAMPLES

**Example 1**: The following example shows that in some pathological situations it is indeed possible to complete a partial matrix in such a way that the completion is symmetric and
positive definite and has a larger number of vanishing entries than the Dempster completion. Consider the matrix

\[
M(x) = \begin{pmatrix}
120 & 4 & -15 & x \\
4 & 121 & -1 & 63 \\
-15 & -1 & 118 & 2 \\
x & 63 & 2 & 120
\end{pmatrix}
\]

Dempster’s completion corresponds to \( x = x_d = -(79/5527) \). The associated matrix \( M(x_d) \) has the following inverse:

\[
M(x_d)^{-1} = \begin{pmatrix}
63 & -1 & 1 & 0 \\
-1 & 1000 & -63 & 2 \\
1 & -63 & 4000 & -18 \\
0 & 2 & -18 & 63
\end{pmatrix}
\]

If, on the other hand, we pick \( x = x_m = -(16/929) \), the associated matrix \( M(x_m) \) has the following inverse:

\[
M(x_m)^{-1} = \begin{pmatrix}
8 & 0 & 1 & 1 \\
0 & 8 & 0 & 2 \\
1 & 0 & 8 & 0 \\
1 & 2 & 0 & 8
\end{pmatrix}
\]

which has six vanishing entries. Notice that, as for the Dempster’s Completion, this extension is symmetric and positive definite.

**Example 2:** Let

\[
M(x) = \begin{pmatrix}
5 & 2 & x \\
2 & 5 & 2 \\
w & 2 & 5 \\
1 & z & 2
\end{pmatrix},
\]

Notice that the specified entries of \( M(x) \) are compatible with symmetry and the Toeplitz structure. By extremizing \( \det[M(x)] \), we obtain seven real matrices completing \( M(x) \) to a nonsingular matrix (they are explicitly given at the bottom of the page).

All of these completions have inverse with zeros in positions \((1,3),(2,4),(3,1)\) and \((4,2)\). Notice that only \( M(x_1) \), \( M(x_3) \), and \( M(x_5) \) are symmetric and have a Toeplitz structure. Among these, only \( M(x_3) \) is also positive definite \( (M(x_3) \text{ is indeed the Dempster Completion}) \). \( M(x_2) \), \( M(x_4) \) are symmetric but do not have a Toeplitz structure and \( M(x_6), M(x_7) \) have a Toeplitz structure but they are not symmetric.

We observe that all of the five symmetric completions are also solutions of the problem of extremizing \( \det \begin{pmatrix}
5 & 2 & x & 1 \\
2 & 5 & y & 2 \\
1 & y & 2 & 5
\end{pmatrix} \).

Similarly, all of the five Toeplitz completions are also solutions of the problem of extremizing \( \det \begin{pmatrix}
5 & 2 & x \\
2 & 5 & y \\
1 & y & 2
\end{pmatrix} \).

This example shows that even if the constraints are compatible with symmetry or other matrix properties, there may exist extremizing completions that do not preserve these features.

**VI. THE CASE OF RECTANGULAR MATRICES**

Next, we extend the results obtained in the previous sections to the general case of possibly nonsquare matrices \( M(x) \in \mathbb{R}^{n \times p} \). We assume that \( p \geq n \) (the case \( p < n \) can be dealt with in a dual fashion). As before, let \( \mathcal{I} \subset \{1,2,\ldots,n\} \times \{1,2,\ldots,p\} \), and \( \mathcal{I}^c \) be the complementary subset. To each \((i,j) \in \mathcal{I}\) we associate the unknown \( x_{ij} \). Let \( x \) be the \( k \)-dimensional vector obtained by stacking the \( x_{ij} \) one on top of the other. Define as before a **partial matrix** to be a parametric family of matrices \( M(x) \) whose entries \( M(x)_{ij} = m_{ij} \) are specified for \((i,j) \in \mathcal{I}\), while \( M(x)_{ij} = x_{ij} \) for \((i,j) \in \mathcal{I}^c \). Again, \( m_{ij} \)
and \(x_{ij}\) take real values. Consider the following matrix completion problems:

**Problem 3:** Find full row rank completions \(M(x_0)\) such that the corresponding Moore–Penrose pseudoinverse \(M(x_0)^\dagger\) satisfies \([M(x_0)]_{i,j} = 0\) for all \((i,j) \in I^\top\).

Notice that, since \(M(x_0)\) is full row rank, the Moore–Penrose pseudoinverse is also a right-inverse and is explicitly given by

\[
M(x_0)^\dagger = M(x_0)^\top (M(x_0)M(x_0)^\top)^{-1}. \tag{20}
\]

**Problem 4:** Find full row rank completions \(M(x_0)\) that extremize \(
\det \left[ \left( M(x)M(x)^\top \right)^{1/2} \right].
\)

**Remark:** The form of the index in Problem 4 is inspired by the fact, established in Lemmata 1 and 2, that in the square problem the variational analysis only depends on the positive part \(P(M(x)) = [M(x)M(x)^\top]^{1/2}\) of \(M(x)\). Actually, only the singular values of \(M(x)\) come into play.

The main result of this section consists in showing that Problem 3 and 4 are equivalent.

**Theorem 3:** \(M(x_0)\) solves Problem 3 if and only if it solves Problem 4.

We first notice that Problem 4 is equivalent to the following: Find a full row rank completion \(M(x_0)\) that extremizes

\[
J(M(x)) := \log \det \left[ \left( M(x)M(x)^\top \right)^{1/2} \right], \tag{21}
\]

Denote by \(D[J(M);\delta M]\) the directional derivative of \(J\) in direction \(\delta M \in \mathbb{R}^{m \times p}\):

\[
D[J(M);\delta M] = \lim_{\varepsilon \to 0} \frac{J(M + \varepsilon \delta M) - J(M)}{\varepsilon}. \tag{22}
\]

We have

**Lemma 4:** If \(M\) is full row rank then, for any \(\delta M \in \mathbb{R}^{m \times p}\),

\[
D[J(M);\delta M] = \text{tr} [M^\dagger \delta M]. \tag{23}
\]

**Proof:** We first observe that if \(Q = Q^\top > 0\), the following expression holds [6], [20], Lemma 2:

\[
D[\log \det (Q);\delta Q] = \text{tr} [Q^{-1}\delta Q]. \tag{24}
\]

Now let

\[
P_M := (MM^\top)^{1/2}, \quad U_M := P_M M \tag{25}
\]

so that

\[
J(M) = \log \det (P_M), \tag{26}
\]

Observe that \(M = P_M U_M\) is the generalized polar decomposition of \(M\). In particular, \(U_M\) is a matrix whose rows are orthonormal: \(U_M U_M^\top = I\). From (20), we get the following representation for the Moore–Penrose pseudoinverse of \(M\)

\[
M^\dagger = U_M^\top P_M^{-1}. \tag{27}
\]

Employing the chain rule, from (26) and (24), we readily get

\[
D[J(M);\delta M] = D[\log \det (P_M);\delta M] D[P_M;\delta M] \tag{28}
\]

In view of (25), the directional derivative \(D[P_M;\delta M]\) may be expressed as

\[
D[P_M;\delta M] = D[MU_M^\top;\delta M] = \delta MU_M^\top + MD[U_M^\top;\delta M]. \tag{29}
\]

From (28) and (29), using the cyclic property of the trace and (27), we now get

\[
D[J(M);\delta M] = \text{tr} [P_M^{-1}\delta MU_M^\top] + \text{tr} [P_M^{-1} MD[U_M^\top;\delta M]] = \text{tr} [M^\dagger \delta M] + \text{tr} [U_M D[U_M^\top;\delta M]]. \tag{30}
\]

The result now follows by observing that \(\text{tr} [U_M D[U_M^\top;\delta M]] = 0\). Indeed,

\[
0 = \text{tr} [U_M D[U_M^\top;\delta M]] = \text{tr} [U_M D[U_M^\top;\delta M]] + \text{tr} [D[U_M;\delta M] U_M^\top] = 2 \text{tr} [U_M D[U_M^\top;\delta M]]. \tag{31}
\]

The following result is a simple generalization of Lemma 3.

**Lemma 5:** Consider the space \(\mathbb{R}^{m \times p}\) endowed with the inner product \(M_1, M_2 := \text{tr} [M_1^\top M_2]\). Let \(M\) be the subspace of \(\mathbb{R}^{m \times p}\) consisting of the matrices whose entries in position \((i,j) \in I\) are zero. Let \(M \in M^\perp\). Then, \([M]_{i,j} = 0\) for all \((i,j) \in I\).

**Proof of Theorem 3:** Problem 4 is equivalent to the following variational problem

extremize \(\{J(M) : e_i^\top Me_j = m_{ij}, (i,j) \in I\}\) \tag{32}

where \(J\) is given by (21). The corresponding Lagrangian is

\[
\mathcal{L}(M) = J(M) + \sum_{(i,j) \in I} \lambda_{i,j}(m_{ij} - c^\top Me_j) \tag{33}
\]

whose unconstrained extremization is obtained by annihilating the directional derivative of \(\mathcal{L}\) in any direction \(\delta M\). In view of Lemma 4, we get

\[
\text{tr} \left[ \left( M^\dagger - \sum_{(i,j) \in I} \lambda_{i,j} e_i e_j^\top \right) \delta M \right] = 0, \quad \forall \delta M \in \mathbb{R}^{m \times p} \tag{34}
\]

or, equivalently

\[
M^\dagger = \sum_{(i,j) \in I} \lambda_{i,j} e_i e_j^\top \tag{35}
\]

It follows, in particular, that the Moore–Penrose pseudoinverse of any full row-rank critical point \(M\) of the Lagrangian (14) has the form of the index in Problem 4.
zeros in positions $(i,j) \in \mathcal{I}^T$. Moreover, if we can find $\lambda_{ij}^0$ such that the matrix $\sum_{(i,j)\in\mathcal{I}} \lambda_{ij}^0 e_i e_j^T$ is full column-rank and

$$M^0 := \left( \sum_{(i,j)\in\mathcal{I}} \lambda_{ij}^0 e_i e_j^T \right)$$

satisfies

$$e_i^T M^0 e_j = m_{ij}, \quad (i,j) \in \mathcal{I} \quad (37)$$

then $M^0$ is indeed a solution of Problem 4.

Assume now that $M(x_0)$ solves Problem 3. Then, $[M(x_0)]$ is full row-rank and $[M(x_0)]^T$ has the form (35). Moreover, $M(x_0)$ satisfies (37). Hence, it solves Problem 4.

Conversely, let $M(x_0)$ solve Problem 4. This is equivalent to

$$D[J(M),\delta M]\big|_{M=M(x_0)} = \text{tr}[M(x_0)\delta M] = ([M(x_0)]^T, \delta M) = 0, \quad \forall \delta M \in \mathcal{M} \quad (38)$$

where $\mathcal{M}$ is the subspace of $\mathbb{R}^{n \times p}$ (defined in Lemma 5) consisting of matrices whose entries in position $(i,j) \in \mathcal{I}$ are zero. Thus, $[M(x_0)]^T \in \mathcal{M}$. By Lemma 5, $[M(x_0)]_{k,j} = 0$ for all $(i,j) \in \mathcal{I}$, namely $M(x_0)$ solves Problem 3.

We outline again the significance of the above result for the solution of systems of linear equations. Consider

$$XM = B \quad (39)$$

where $B$ is given. Suppose that only certain elements of $M$ could be estimated/determined. If the completion $M_0$ has full row rank, we can associate to it the solution

$$X_{M_0} = BM_0^\dagger, \quad (40)$$

Again we identify as desirable, according to the principle of parsimony, completions $M_0$ such that $M_0^\dagger$ has a maximum number of zero entries.

VII. DISCUSSION

The nuclear norm (sum of singular values) of a matrix is often used in convex heuristics for rank minimization problems in control, signal processing, and statistics. It has been employed in a series of recent papers on matrix completion, see [10], [11], [26], [31] and references therein. The renewed interest in this metric has both theoretical and practical reasons as argued in the above mentioned papers. Variational problems involving the sum of the logarithm of the singular values (the logarithm of the determinant in the covariance case), such as those presented in this paper, occupy a somewhat complementary place. Indeed, as we have shown above, they lead to constraints on the (pseudo)inverse of $M$, while in the latter papers a low rank matrix is sought. Moreover, in the case when $M = \Sigma$ is a covariance matrix, Dempster’s Completion $\Sigma^0$ maximizes entropy, namely the sum of the logarithm of the singular values (eigenvalues), whereas in [10], [11], [26], [31] the sum of the singular values is minimized.

Our variational problems appear close in spirit to Janes [24], [25], where, following in the footsteps of Boltzmann (1877), Schrödinger [34], Cramér [14], Sanov [33], etc., and followed by such corphaei as Dempster himself [16], Akaike [2], Burg [5], [6], etc., he promoted maximum entropy methods to general inference methods. It might actually be worthwhile to quote an illuminating passage from the introduction of [25] which deals with spectral analysis: “There are many different spectral analysis problems, corresponding to different kinds of prior information about the phenomenon being observed, different kinds of data, different kinds of perturbing noise, and different objectives. It is, therefore, quite meaningless to pass judgment on the merits of any proposed method unless one specifies clearly: ‘In what class of problems is this method intended to be used?’” Most of the current confusion on these questions is, in the writer’s opinion, the direct result of failure to define the problem explicitly enough.” We feel that these considerations apply equally well to the matrix completion problem.

In this paper, we have chosen to discuss a very general completion problem where no further requirement is imposed on the solution matrix. As soon as the solution is required to feature some properties, such as being positive definite, (with the possible additional constraints of being Toeplitz, circulant, etc.) the existence of matrices having the prescribed elements and properties becomes an issue. When existence is guaranteed, it should be apparent that our variational analysis can be readily adapted to these more structured problems.

We finally want to observe that completing a matrix so that it enjoys certain properties [4], [6], [10]–[12], [15], [16], [22], [23], [26], [28], [29], [31], [35] may be viewed as a generalized moment problem. These are problems where a function (a measure, a matrix, etc.) is sought satisfying certain given moment constraints as in the classical moment problem [3], [27], but also enjoying further properties. These may take several different forms. We mention the important case of bounds on the complexity, such as a bound on the degree of the rational solution, for applications in communications and control engineering, cf. e.g., [7]–[9], [17]–[20], [32].

ACKNOWLEDGMENT

The authors wish to thank two anonymous reviewers for a number of useful hints and insightful comments. In particular, one of them suggested that a simpler proof of Lemma 2 could be established.

REFERENCES

Augusto Ferrante was born in Prove di Sacco, Italy, on August 5, 1967. He received the “Laurea” degree (cum laude) in electrical engineering in 1981 and the Ph.D. degree in control systems engineering in 1995, both from the University of Padova, Italy.

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.

Michele Pavon was born in Venice, Italy, on October 12, 1950. He received the Laurea degree from the University of Padova, Padova, Italy, in 1974, and the Ph.D. degree from the University of Kentucky, Lexington, in 1979, both in mathematics.

After service in the Italian Army, he was on the research staff of LADSEB-CNR, Padova, Italy, for six years. Since July 1986, he has been a Professor at the College of Engineering, the University of Padova. He has visited several institutions in Europe and North America. His present research interests include maximum entropy problems and quantum control.