JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS **102**, 30–37 (1984)

A Constrained Matrix Factorization Problem*

ROMANO M. DESANTIS

Department of Electrical Engineering, Ecole Polytechnique de Montreal, Montreal, Quebec H3C 3A7, Canada,

AND

WILLIAM A. PORTER

Department of Electrical Engineering, Louisiana State University, Baton Rouge, Louisiana 70803

Submitted by A. Schumitzky

The present paper considers a constrained matrix factorization problem which is a natural generalization of the classical triangular factorization.

INTRODUCTION

Denoting with $W \triangleq [w_{ij}]$ and $\Gamma \triangleq [\gamma_{ij}]$ two $n \times n$ matrices, let S be a subset of $\{1,..., n\}^2$. W is *S-Constrained* if $w_{ij} = 0$ for $ij \notin S$; Γ is $S^\perp$-*Constrained* if $\gamma_{ij} = 0$ for $ij \in S$. The objective of the present paper is to state, prove and discuss the solution to the following *S-Constrained Matrix Factorization Problem*: given an $n \times n$ matrix A and a subset $S \subset \{1,..., n\}^2$, determine S-Constrained W and $S^\perp$-Constrained Γ so that

$$A = (I + W)^{-L}(I + \Gamma) \tag{1}$$

where I is the identity matrix and $(I + W)^{-L}$ is the left inverse of $I + W$.

Of a more general type than those usually considered in the technical literature, this factorization is of interest because (as will be illustrated in the sequel) it represents a key step in the solution of a number of digital signal and image processing problems. Also, by specializing the choice of the set S, it can be made to coincide with the familiar Schur Coleski triangular factorizations [2] as well as with the more recent angular factorizations considered in [4].

* Sponsored in part under NSF Grant 78/88/71, AFOSR Grant 78-3500 and Canadian Research Council Grant CNRC-A-8244.

0022-247X/84 $3.00

THE MAIN RESULT

Our main result is embodied in the following theorem.

THEOREM. *An $n \times n$ matrix $A = I + Q$ always admits an S-constrained matrix factorization. One such factorization is characterized by*

$$W_0 = -\sum_{i=1}^{n} \Delta_i Q P^i (I + P^i Q P^i)^{-L} \tag{2}$$

where the matrices Δ_i and P^i are defined as follows: $\Delta_i \triangleq [\delta_{nm}]$, $\delta_{nm} = 0$ for $nm \neq ii$, $=1$ for $nm = ii$; $P^i \triangleq \sum_{j \in S_i} \Delta_j$, $S_i \triangleq \{j \in S_i \mid ij \in S\}$; $(I + P^i Q P^i)^{-L}$ denotes the left inverse of $I + P^i Q P^i$. A necessary and sufficient condition for this factorization to be unique is that $I + P^i Q P^i$ be invertible for each $i \in \{1,..., n\}$.

COROLLARY 1 [2. Theorem 1, p. 20]. *For a given matrix $A \triangleq [a_{ij}]$, $ij \in \{1,..., n\}^2$, let the leading submatrices A_i, $i \in \{1,..., n\}$ be nonsingular ($A_i \triangleq P^i A P^i$; $P^i \triangleq \sum_{j=1}^{i} \Delta_j$; $\Delta_i \triangleq [\delta_{sl}]$, $\delta_{sl} = 0$ for $sl \neq ii$, $=1$ for $sl = ii$). Then A may be represented by the product*

$$A = (I + W)\, D(I + V)$$

where

W is lower triangular ($W \triangleq [w_{ij}]$, $w_{ij} = 0$ for $j \geqslant i$);
V is upper triangular ($V \triangleq [v_{ij}]$, $v_{ij} = 0$ for $i \geqslant j$);
D is zero lower/upper triangular ($D \triangleq [d_{ij}]$, $d_{ij} = 0$ for $i \neq j$).

COROLLARY 2 [4, Theorem 2]. *For a given array $A \triangleq [a_{ijsl}]$, $ijsl \in \{1,..., n\}^4$, let the leading subarrays A_{ij}, $ij \in \{1,..., n\}^2$, be nonsingular ($A_{ij} \triangleq P^{ij} A P^{ij}$; $P^{ij} \triangleq \sum_{s=1}^{i} \sum_{l=1}^{j} \Delta_{sl} - \Delta_{ij}$; $\Delta_{ij} \triangleq [\delta_{slnm}]$, $\delta_{slnm} = 0$ for $sl \neq nm \neq ij$, $=1$ for $sl = nm = ij$). Then A may be represented by the product*

$$A = (I + W)\, D(I + V)$$

where

W is lower quadrangular ($W \triangleq [w_{ijnm}]$, $w_{ijnm} = 0$ for: $ij = nm$; $i > n$ or $j > m$);

V is upper quadrangular ($V \triangleq [v_{ijnm}]$, $v_{ijnm} = 0$ for: $ij = nm$, $i < n$ or $j < m$);

D is zero lower/upper quadrangular ($D \triangleq [d_{ijnm}]$, $d_{ijnm} \neq 0$ *only for*: $ij = nm$; $i > n$ *or* $j > m$; $i < n$ *or* $j < m$).

EXAMPLE. Consider a 4×4 matrix

$$Q = \begin{bmatrix} q_{11} & q_{12} & q_{13} & q_{14} \\ q_{21} & q_{22} & q_{23} & q_{24} \\ q_{31} & q_{32} & q_{33} & q_{34} \\ q_{41} & q_{42} & q_{43} & q_{44} \end{bmatrix}$$

and let $S = \{11, 12, 31\}$. The theorem says that the matrix $I + Q$ has the representation

$$I + Q = (I + W)^{-\mathrm{L}}(I + \Gamma)$$

with W and Γ, respectively, of the form

$$W = \begin{bmatrix} X & X & 0 & 0 \\ 0 & 0 & 0 & 0 \\ X & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\Gamma = \begin{bmatrix} 0 & 0 & X & X \\ X & X & X & X \\ 0 & X & X & X \\ X & X & X & X \end{bmatrix}$$

W and Γ can be computed by recognizing that in this particular case one has

$$\Delta_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \qquad \Delta_2 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\Delta_3 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \qquad \Delta_4 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$S_1 = \{1, 2\}, \qquad S_2 = 0, \qquad S_3 = \{1\}, \qquad S_4 = 0.$$

Hence, $P^2 = P^4 = 0$, and

$$P^1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \qquad P^3 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

It follows (using Eq. (2))

$$W_0 = \begin{bmatrix} W_{11} & W_{12} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ W_{31} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

with

$$W_{11} = \frac{(1 + q_{22})\, q_{11} - q_{12} q_{21}}{(1 + q_{11})(1 + q_{22}) - q_{12} q_{21}}$$

$$W_{12} = \frac{q_{12}}{(1 + q_{11})(1 + q_{22}) - q_{12} q_{21}}$$

$$W_{31} = \frac{q_{31}}{1 + q_{11}}.$$

Using Eq. (1) one has $\Gamma = -Q + W(I + Q)$. It follows

$$\Gamma = \begin{bmatrix} 0 & 0 & \Gamma_{13} & \Gamma_{14} \\ \Gamma_{21} & \Gamma_{22} & \Gamma_{23} & \Gamma_{24} \\ 0 & \Gamma_{32} & \Gamma_{33} & \Gamma_{34} \\ \Gamma_{41} & \Gamma_{42} & \Gamma_{43} & \Gamma_{44} \end{bmatrix}$$

with

$$\Gamma_{13} = q_{13}(W_{11} - 1) + W_{12} q_{23}$$

$$\Gamma_{14} = q_{14}(W_{11} - 1) + W_{12} q_{24}$$

$$\Gamma_{21} = -q_{21}; \qquad \Gamma_{22} = -q_{22}; \qquad \Gamma_{23} = -q_{23}; \qquad \Gamma_{24} = -q_{24}$$

$$\Gamma_{32} = -q_{32} + W_{32} q_{12}; \qquad \Gamma_{33} = -q_{33} + W_3 + q_{13}$$

$$\Gamma_{34} = -q_{34} + W_{31} q_{14}$$

$$\Gamma_{41} = -q_{41}; \qquad \Gamma_{42} = -q_{42}; \qquad \Gamma_{43} = -q_{34}; \qquad \Gamma_{44} = -q_{44}.$$

By selecting a different set S a variety of other structure forms for the matrices W and Γ can be considered.

Proof of the Theorem. Observe that the matrix $W \triangleq [\omega_{ij}]$ is such that $w_{ij} = 0$ for $ij \notin S$ if and only if for every vector pair $xy \in R^n$ such that $x = Wy$ one has

$$x = \sum_{i=1}^{n} \Delta_i W \sum_{j=1}^{n} \Delta_j y = \sum_{i=1}^{n} \Delta_i W \sum_{j \in S_i}^{n} \Delta_j y.$$

Clearly then, a necessary and sufficient condition for W to enjoy the property $w_{ij} = 0$ for $ij \notin S$ is that W have the representation $W = \sum_{i=1}^{n} \Delta_i WP^i$ (the notations are defined in the statement of the theorem). In the same vein, a necessary and sufficient condition for Γ to be such that $\gamma_{ij} = 0$ for $ij \in S$ is $\sum_{i=1}^{n} \Delta_i \Gamma P^i = 0$.

To prove the first part of the theorem one then simply needs to show that

$$W_0 = \sum_{i=1}^{n} \Delta_i W_0 P^i$$

and

$$\sum_{i=1}^{n} \Delta_i [(I + W_0)(I + Q) - I] P^i = 0.$$

The first of these two properties follows by observing that

$$P^i [I + P^i Q P^i]^{-\mathrm{L}} = P^i [I + P^i Q P^i]^{-\mathrm{L}} P^i.$$

The second property is proved by a direct inspection

$$\begin{aligned}
&\Delta_i (I + W_0)(I + Q) P^i - \Delta_i P^i \\
&\quad = \Delta_i \left(I - \sum_{i=1}^{n} \Delta_i Q P^i [I + P^i Q P^i]^{-\mathrm{L}} \right) (I + Q) P^i - \Delta_i P^i \\
&\quad = \Delta_i P^i + \Delta_i Q P^i - \Delta_i Q P^i [I + P^i Q P^i]^{-\mathrm{L}} (I + P^i Q P^i) P^i - \Delta_i P^i \\
&\quad = \Delta_i Q P^i - \Delta_i Q P^i = 0.
\end{aligned}$$

The uniqueness part of the theorem is obtained by verifying, again by a direct inspection, that under the hypothesis of the invertibility of $I + P^i Q P^i$, for each $i \in \{1, \ldots, n\}$, the properties $W = \sum_{i=1}^{n} \Delta_i W P^i$ and $\sum_{i=1}^{n} \Delta_i [(I + W)(I + Q) - I] P^i = 0$ imply $W = W_0$.

Application to Digital Image Processing

Let the luminosity of a digital image be represented in terms of a family of random variables $x(ij)$ indexed by $i \in \{1,..., p\}$, $j \in \{1,..., m\}$ and characterized by $E[x(ij)] = 0$ and $E[x(ij)\, x(sl)] = q_{ijsl}$ (the symbol $E[r]$ denotes the expected value of r). Let $\eta(sl)$ denote a family of zero mean value random variables independent from $x(ij)$ and such that

$$E[\eta(ij)\, \eta(sl)] = 1 \qquad \text{for} \quad ij = sl, \qquad =0 \qquad \text{for} \quad ij \neq sl;$$

let $z(ij) = x(ij) + \eta(ij)$ describe the image corrupted by additive noise data. Consider the class of window constrained linear filters, W, described by

$$\hat{x} = Wz$$

implies

$$\hat{x}(ij) = \sum_{s=i-N_2}^{i+N_1} \sum_{p=j-M_2}^{i+M_1} w_{ijsl}\, z(sl)$$

(N_1, N_2, M_1, M_2 assigned positive integers). Consider the problem of determining, in this class, the filter which minimizes

$$E \left| \sum_{ij} (\hat{x}(ij) - x(ij))^2 \right| \triangleq J(W).$$

General background material regarding this problem can be found in [5]. Following this reference, we represent the image x as given by a random $n \triangleq p \times m$-dimensional vector (column scanning representation)

$$x^t = [x(00) \cdots x(0m)\, x(10) \cdots x(1m) \cdots x(nm)]^T.$$

This random vector is characterized by a zero mean value and an $n \times m$ covariance matrix

$$Q_x \triangleq [q_{ij}] \triangleq \begin{bmatrix} q_{0000} & \cdots & q_{00nm} \\ \vdots & & \vdots \\ q_{nm00} & \cdots & q_{nmnm} \end{bmatrix}$$

The image corrupted by noise data is also represented by a zero mean valued random vector

$$z = [z_{00} \cdots z_{nm}]^T$$

with correlation matrix $Q_z = I + Q_x$.

A filter with the desired constraint is then represented in terms of the $n \times n$ matrix

$$W \triangleq \begin{bmatrix} w_{0000} & \cdots & w_{00nm} \\ \vdots & & \vdots \\ w_{nm00} & \cdots & w_{nmnm} \end{bmatrix}$$

with entries subject to $w_{ijs} = 0$ for $ijsl \notin S$, where

$$S \triangleq \{ijsl \in S \mid -N_2 < (s-i) < N_1, -M_2 < l-j < M_1\}.$$

Observing that

$$J(W) = \operatorname{tr}\{WQ_x W' + WW' - WQ_x - Q_x W' + Q_x\}$$

where W' denotes the transpose of W, it follows that given any two admissable filters W_0 and W one has

$$J(W) - J(W_0) = \operatorname{tr}\{[(W_0 + \Delta W)(I + Q_x) - Q_x] \Delta W'\}$$

where $\Delta W \triangleq W - W_0$. It follows that a necessary and sufficient condition for W_0 to be optimal is

$$\operatorname{tr}\{[W_0(I + Q_x) - Q_x] \Delta W\} = 0.$$

Clearly this condition is satisfied if and only if

$$W_0(I + Q_x) - Q_x \triangleq \Gamma \triangleq [\gamma_{ijsl}]$$

is such that $\gamma_{ijsl} = 0$ for $ijsl \in S$, which is equivalent to

$$(I - W_0)(I + Q_x) = I + \Gamma$$

where $W \triangleq [\omega_{ijsl}]$ is such that $w_{ijsl} = 0$ for $ijsl \notin S$ and $\Gamma \triangleq [\gamma_{ijsl}]$ is such that $\gamma_{ijsl} = 0$ for $ijsl \in S$.

One can at this point apply Theorem 1, observing that $I + Q_x$ is positive definite and clearly $I + P^i Q_x P^i$ is invertible. Theorem 1 says then that the optimal filter exists, is unique and is given by Eq. (2).

The above development holds "verbatim" if instead of considering a window constrained filter we had considered other types of constraints such as column processing, row processing or row and column processing filters (see [5, Chapter 8]). The difference in the consideration of one type of constraint over the other is merely in the appropriate choice of S.

References

1. R. M. DeSantis and W. A. Porter, Optimization problems in partially ordered Hilbert resolution space, *Internat. J. Control*, in press.
2. U. N. Faddeeva, "Computational Methods of Linear Algebra," Dover, New Yor, 1959.
3. I. Z. Gohberg and M. C. Krein, "Theory of Volterra Operators in Hilbert Space and Applications," Vol. 18, Amer Math Soc., Providence, R. I., 1969.
4. W. A. Porter and R. M. DeSantis, Angular factorization of matrices, *J. Math. Anal. Appl.* **88** (1982), 591–603.
5. W. K. Pratt, "Digital Image Processing," Wiley, New York, 1978.