Confluent Vandermonde matrices using Sylvester's structure

Lamine Melkemi

Laboratoire de l'Informatique du Parallélisme

École Normale Supérieure de Lyon
Unité de recherche associée au CNRS n° 1398

Confluent Vandermonde matrices using Sylvester's structures

Lamine Melkemi

March 1998

Research Report N° 98-16

École Normale Supérieure de Lyon
46 Allée d'Italie, 69364 Lyon Cedex 07, France
Téléphone : +33(0)4.72.72.80.00
Télécopieur : +33(0)4.72.72.80.80
Adresse électronique : `lip@ens-lyon.fr`

Confluent Vandermonde matrices using Sylvester's structures

Lamine Melkemi

March 1998

Abstract

In this paper, we first show that a confluent Vandermonde matrix may be viewed as composed of some rows of a certain block Vandermonde matrix. As a result, we derive a Sylvester's structure for this class of matrices that appears as a natural generalization of the straightforward one known for usual Vandermonde matrices. Then we present some applications as an illustration of the established structure. For example, we show how confluent Vandermonde and Hankel matrices are linked with each other, and also we describe an $\mathrm{O}(n^2)$ algorithm for solving confluent Vandermonde least squares minimizations problems.

Keywords: confluent Vandermonde matrix, Sylvester's equation, structured matrices.

Résumé

Dans cet article, nous proposons une structure de Sylvester pour les matrices de Vandermonde confluentes qui parait comme une généralisation naturelle de celle connue dans le cas d'un système de Vandermonde simple. La démonstration de ce résultat tire profit d'une propriété intéressante disant, dans un sens que nous préciserons plus loin, qu'une telle matrice est en fait plongée dans une matrice de Vandermonde par blocks. En exploitant cette structure, nous montrons ensuite, qu'il est possible d'exprimer l'inverse d'une matrice de Vandermonde confluente comme le produit de son duale et d'une matrice de Hankel. Toujours à l'aide de cette structure, nous décrivons un algorithme $\mathrm{O}(n^2)$ permettant de résoudre les problèmes aux moindres carrés basés sur ces matrices. Enfin nous montrons comment on peut étendre les résultats établis à une classe de matrices beaucoup plus génerale.

Mots-clés: matrice de Vandermonde confluente, structure de Sylvester, matrices structurees.

1. Introduction. Given n different numbers $x_0,\ x_1,\ \ldots, x_{n-1}$ (that is $x_i \neq x_j$ for $i \neq j$). We mean by a confluent Vandermonde matrix the following $m \times (p+1)n$ matrix:

$$V_{p+1} = [f(x_0)\ \ f'(x_0)\ \ \ldots\ \ f^{(p)}(x_0)\ \ \ldots\ \ f(x_{n-1})\ \ f'(x_{n-1})\ \ \ldots\ \ f^{(p)}(x_{n-1})]$$

where $f(x)$ is the vectorial function:

$$f(x) = (1, x, x^2, \ldots, x^{m-1})^t$$

and $f'(x)$,$f^{(2)}(x)$,..., $f^{(p)}(x)$ are the derivatives of $f(x)$. In fact, the number of the derivatives p should vary as x_i, but for convenience and without loss in generality, it is assumed throughout the paper, that it is fixed. It is the matter of a well known concept that naturally generalizes usual Vandermonde systems and can be viewed as a matrix representation of some interpolation problems [11]. For example, Hermite's interpolation amounts simply to solve the system of linear equations: $V_p^T x = b$. Because of its importance in applications (see [2] for instance), such matrices have received a particular attention in the literature. In particular, several $\mathrm{O}(n^2)$ algorithms have been proposed for computing the inverse of Vandermonde and confluent Vandermonde matrices and that of their duals ([4], [5], [20]). In [13], Higham considered the more general case where the monomial x^k is replaced by a polynomial $P_k(x)$ and constructed what is called a confluent Vandermonde-like matrix. Then he developed $\mathrm{O}(n^2)$ solutions for computing the inverse of such a matrix for particular polynomials. In [14], he analyzed the numerical stability of these methods. More recently, Lu [17], [18] showed the ability to design asymptotically faster solutions to these problems and conceived in particular an algorithm with $\mathrm{O}(np(\log np)\log n)$ running time for inverting $np \times np$ confluent Vandermonde matrices. His approach is based on fast convolution products and fast polynomial divisions [1].

On the other side, it is common to put matrices with particular forms, such as Toeplitz matrices, into a unified algebraic formalism. In other words, given two "simple" matrices L and U the idea is to find the Sylvester's equation :

$$LM - MU = GH$$

where the concerned matrix M is completely characterized by the generator GH which is, in the essence, a summation of a fixed number of outer products. The importance of this equation called the structure of M stems from the fact that it constitutes a powerful tool for efficiently manipulating such matrices, as it is emphasized in different situations. Of course, the structure is more and more elegant as long as the matrices L and U are simple and GH is of a small rank. See [15] and the survey paper [16] with the references therein.

In this paper, we consider this aspect of manipulating confluent Vandermonde matrices via their Sylvester's structure which is deduced from a useful observation that a confluent Vandermonde matrix is in a certain viewpoint embedded into a block-Vandermonde one. After some notations in section 2, and a brief discussion about block-Vandermonde matrices in section 3, where we show some limitations when trying to extend the concepts available in the usual case and evidence particular instances allowing to palliate the difficulties, we present in Section 4 our main result where a Sylvester's equation satisfied by the confluent Vandermonde matrices is established and proved. Examining it, one may assert that it well characterizes the confluent Vandermonde matrices and, apparently, can not be sharpened further. To our knowledge such a structure does not exist elsewhere. In section 5, we study the converse situation where it is shown that up to a slight modification, it is possible to determine precisely the matrices which are solutions of the modified Sylvester's equation. In other words, we establish an interesting canonical structure to which the studied Sylvester's equation may be easily transformed and conversely. As an application, an inverse formula is given. The sections 6, 7 and 8 may be viewed as interesting applications of the Sylvester's structure. In section 6, a connection between confluent Vandermonde matrices, Hankel and Toeplitz-like ones is shown in a way similar to that found out for the usual Vandermonde matrices (see [7], [10], [12]). Nevertheless, it is shown that though the structure established, there are severe limitations as one attempts a direct generalization. Fortunately, it is possible to compensate the drawbacks so that one may construct $\mathrm{O}(n\log^2 n)$ algorithms for solving weighted least squares minimizations settling overdetermined Hermite's interpolations. In section 7, we develop an $\mathrm{O}(n^2)$ method for computing the Cholesky factorization of the covariance matrix of the confluent Vandermonde matrix. The algorithm as it is well known may be exploited for solving confluent Vandermonde least squares minimizations problems. Finally, we consider in section 8, the notion of generalized confluent Vandermonde matrices and explain how the results available in the simple case may be extended.

2. Notations. To make the description of the text as clear as possible, some notations are adopted. As we will deal with rectangular matrices and block-Vandermonde matrices, it follows that different linear spaces will be considered. Consequently, care must be taken about the use of a same notation.

For this reason, we denote throughout the paper by $(e_i)_{0\le i\le m-1}$, $(f_j)_{0\le j\le p-1}$, $(h_k)_{0\le k\le mp-1}$, and $(g_k)_{0\le k\le mp-1}$, the canonical basis in the linear spaces $\mathbb{C}^m$, $\mathbb{C}^p$, $\mathbb{C}^{mp}$ and $\mathbb{C}^{np}$ respectively, where $\mathbb{C}$ is the complex field. In general, we consider the canonical basis in $\mathcal{R}^m$ where $\mathcal{R}$ is a ring with unit element, and we denote it by $(E_0,\ldots,E_{m-1})$. For example, if $\mathcal{R} = \mathbb{C}^{p\times p}$ is the ring of the $p\times p$ complex matrices then for $0\le k\le m-1$, we have:

$$E_k = [h_{kp}, h_{kp+1},\ldots,h_{(k+1)p-1}]$$

which is a $mp\times p$ matrix. Now, let say that whatever canonical basis one uses, it is systematically understood along the paper that if u is a vector then u_{i-1} denotes its i^{th} coordinate.

Since one of our main objectives is to construct a Sylvester's structure for confluent Vandermonde matrices, it is natural to expect the use the displacement matrix defined as a square matrix with $1'$s along its first sub-diagonal and $0'$s elsewhere. If we have a $k\times k$ displacement matrix, let denote it by Z_k. For simplicity and unless otherwise stated, we let: $Z = Z_m$ and $\bar{Z} = Z_{np}$. The well known reverse identity matrix may be also utilized in such a context and is a square matrix with $1'$s along its anti-diagonal and $0'$s elsewhere. We will denote by J_k a reverse identity matrix of size $k \times k$.

In this paper for reasons to be seen latter, we distinguish between vector transpose and matrix transpose and we use two notations: u^t means the transpose of the vector u whereas R^T means the transpose of the matrix R. By the way, we may use the anti-transpose operator, T_2, defined as follows:

$$A^{T_2} = J_k A^T J_k$$

where A is a $k \times k$ matrix.

3. Block-Vandermonde Matrices. Clearly, the Vandermonde matrices may be defined over arbitrary rings with unit element, as for example the ring of matrices, and it is of interest to note that in this straightforward generalization, the well known structure of Vandermonde matrices remains safe. That is, if $(\mathcal{R}, +, ., 0, 1)$ denotes such a ring, then we have:

$$Z^T V - VD = E_{m-1} U^t$$

where $V = V(B_0, B_1, ..., B_{n-1})$ is the $m \times n$ Vandermonde matrix with elements B_i into the ring $\mathcal{R}$, Z is the displacement matrix (with 1 ($\in \mathcal{R}$) in the first sub diagonal and 0 ($\in \mathcal{R}$) elsewhere), $D = diag(B_i)$ and $U^t = -(B_0^n, \ldots, B_{n-1}^n)$. Recall that E_{m-1} stands for the last element in the canonical basis of $\mathcal{R}^n$.
In the case where $\mathcal{R}$ is the ring of $p \times p$ matrices defined over the field of the complex numbers $\mathbb{C}$, we obtain what is known as block-Vandermonde matrices. Here, the displacement matrix in the above Sylvester's equation must be understood as the p-power Z_{mp}^p of the displacement matrix Z_{mp} with elements in the field $\mathbb{C}$. Based on this structure, it is thus possible to construct efficient algorithms for manipulating block-Vandermonde matrices. On the other hand, there are natural reasons to expect that most operations over Vandermonde matrices may be carried out to block-Vandermonde ones requiring up to a p factor the same amount of computation. Unfortunately, this is not true in general because of the simple fact that several flexible properties such as commutativity are missed in the ring $\mathcal{R}$ while well verified in the field $\mathbb{C}$. Consider for instance the problem of solving a linear system of equations: $V^T x = b$ where V is a block-Vandermonde matrix. An approach due to [7] (see also [10]) is to transform this system to the following equivalent one:

$$VV^T x = Vb.$$

The idea lies in the observation that in the case where V is a Vandermonde matrix, one may assert that VV^T is a Hankel matrix. Hence the complexity for solving the considered system is the same as for solving a Hankel linear system. Since there are $\mathrm{O}(n\log^2 n)$ algorithms for computing the inverse of a positive definite Hankel matrix (see [3], [6], [8], [19] among others), one finds out another way of constructing $\mathrm{O}(n\log^2 n)$ algorithms for the polynomial interpolation. On the contrary, when attempting to extend this technique to block-Vandermonde matrices, the difficulty arises as soon as one writes the structure:

$$V^T Z^p - D^T V^T = U E_{n-1}^t$$

of V^T, where in general, we have $D \neq D^T$ (recall that $D = diag(B_i)$). As an alternative issue, this last remark evidences possible ways for compensating such a disadvantage. Actually, if we assume that the blocks B_i are symmetric matrices then one can check that VV^T is a block Hankel matrix. Indeed, if we apply the matrix V to the above structure of V^T (where now $D = D^T$), we obtain :

$$VV^T Z^p - VDV^T = VUE_{n-1}^t.$$

Using the fact that $VD = Z^{pT}V - E_{n-1}U^t$, the following structure is easily deduced:

$$VV^T Z^p - Z^{pT} VV^T = VUE_{n-1}^t - E_{n-1}U^t V^T$$

which asserts that VV^T is a block-Hankel matrix.

4. Embedding confluent Vandermonde matrices into block Vandermonde ones. In this section, we restrict ourselves to particular block matrices. Given n different complex numbers $x_0, x_1, \dots, x_{n-1}$, we will consider the $p \times p$ matrices $B_0, B_1, ..., B_{n-1}$ defined as follows:

$$B_i = B(x_i) = \begin{bmatrix} x_i & 1 & 0 & \dots & 0 \\ 0 & x_i & 2 & \dots & 0 \\ 0 & 0 & x_i & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & p-1 \\ 0 & 0 & \dots & \dots & x_i \end{bmatrix}$$

Let now form the $mp \times np$ block-Vandermonde matrix $V = V(B_0, B_1, \dots, B_{n-1})$ where for $0 \leq i \leq m-1$ and $0 \leq j \leq n-1$, its (i,j) block-element is B_j^i. As it is previously stated, the matrix V fulfills the Sylvester's equation below:

$$Z_{mp}^{pT} V - VD = E_{m-1} U^t$$

where Z_{mp} is the $mp \times mp$ displacement matrix, D is the block-diagonal matrix such that: $D = diag(B_0, ..., B_{n-1})$, and both U and E_{m-1} are block-vectors defined over $\mathcal{R}^n$ *and* $\mathcal{R}^m$ ($\mathcal{R}$ being the ring of the $p \times p$ complex matrices). It is easily seen that E_{m-1} is the last element in the canonical basis of $\mathcal{R}^m$ and $U^t = -(B_0^m, \dots, B_{n-1}^m)$. The structure of V just evidenced will be a corner stone in constructing that of confluent Vandermonde matrices. More precisely, we will show that the matrix V contains, in a certain sense, a confluent Vandermonde one, and conversely every confluent Vandermonde matrix is embedded into a block Vandermonde one of a same kind. This result, from which the desired Sylvester's structure is obviously derived, is revealed by an interesting observation in the behavior of the powers $B^k(x)$ $(k \geq 0)$ of the matrix $B(x)$.
Indeed, an investigation in the elements of the k-power matrix $B^k(x)$ permits to observe that its first row is interestingly formed by x^k and its derivatives. In general, it is possible to show by an inductive reasoning about the power k, that the (i,j) element α_{ij}^k of $B^k(x)$ can be expressed as follows (we assume for convenience $0! = 1$) :

$$\alpha_{ij}^k = \frac{j!}{i!} \binom{k}{j-i} x^{k-j+i} = \frac{j!k!}{i!(j-i)!(k-j+i)!} x^{k-j+i}$$

for $0 \leq i \leq j \leq i+k$, and is reduced to zero elsewhere. As a consequence, the confluent Vandermonde matrix V_p is simply that formed by the $(kp)^{th}$ rows of V in the natural order $(0 \leq k \leq m-1)$. Mathematically, if P denotes the permutation matrix defined by:

$$P = \sum_{i=0}^{p-1} \sum_{k=0}^{m-1} h_{im+k} h_{kp+i}^t$$

where $(h_s)_{0 \leq s \leq mp-1}$ is the canonical basis of $\mathbb{C}^{mp}$, then we can assert that:

$$PV = \begin{bmatrix} V_p \\ X \end{bmatrix}$$

On the other hand, since $Z_{mp}^{pT} V$ upshifts p times the rows of V so that the $(kp)^{th}$ row of V, $1 \leq k \leq m-1$, will be the $((k-1)p)^{th}$ one of $Z_{mp}^{pT} V$, the following interesting relation is obviously deduced:

$$P Z_{mp}^{pT} V = \begin{bmatrix} Z^T V_p \\ X' \end{bmatrix}$$

In this direction, it is implicitly said that one is applying the permutation matrix P to the structure of V. Hence, it is useful to analyze the aspect of the generator $PE_{m-1}U^t$. For direct matrix manipulations lead us to consider PE_{m-1} as a block-vector, where its j^{th} block-component $(0 \leq j \leq p-1)$ is precisely the $m \times p$ matrix $e_{m-1} f_j^t$. Consequently, the first m rows of $PE_{m-1}U^t$ form the $m \times (np)$ matrix $e_{m-1} y^t = e_{m-1} f_0^t U^t$ where its last row is the first one of U^t and otherwise, the elements are reduced to zero. Clearly $-y^t$ is the first row of $-U^t$ which is simply an alignement of the first rows of the blocks $B_0^m, \ldots, B_{n-1}^m$. It follows from the above observation related to $B^m(x)$ that:

$$-y^t = (x_0^m, m x_0^{m-1}, \ldots, (m)_{p-1} x_0^{m-p+1}, \ldots, x_{n-1}^m, m x_{n-1}^{m-1}, \ldots, (m)_{p-1} x_{n-1}^{m-p+1})$$

where $(m)_k = \frac{m!}{k!}$. Taking into account these considerations and the fact that we are concerned with the first m rows only, as we look forward V_p, we first observe that:

$$P Z_{mp}^{pT} V = P(VD + E_{n-1}U^t) = PVD + PE_{n-1}U^t$$

from which we conclude that the Sylvester's equation of the confluent Vandermonde matrix V_p is given by:

$$Z^T V_p - V_p D = e_{m-1} y^t$$

Clearly, this structure appears as a natural generalization of that of the usual Vandermonde matrices. However, because of the fact that D is a block-diagonal matrix and not a diagonal one, one may expect serious problems, as it will be seen in section 6, when trying to safely carry out, via this equation, the results established in the simple case.

5. Normalization and the inverse formula. In the previous section, we have shown that the confluent Vandermonde matrix V_p is a solution of the Sylvester's equation below:

$$Z^T X - XD = e_{m-1} y^t$$

Conversely, it is possible to determine the matrix X for which this equation is verified. Nevertheless, for arbitrary vectors y, the obtained matrix X has a queer aspect and is not a priori representative in a convinced manner inasmuch one feels being far away from the confluent Vandermonde context. As a remedy, we show here that the vectors y of the form:

$$y^t = (t_0, 0, \dots, 0, t_1, 0, \dots, 0, \dots, t_{n-1}, 0, \dots, 0)$$

give rise to solutions X with elements of simple expressions. Beforehand we suggest a normalization procedure over the Sylvester's structure which allows to conclude that there is no loss in generality when considering the particular y just above.

We first apply over the considered equation a suitable diagonal matrix K so that the non null elements along the super-diagonal of D will be reduced to 1. To this end, let K' and K denote respectively the following $p \times p$ and $np \times np$ diagonal matrices:

$$K' = diag(1, 1!, 2!, \dots, (p-1)!), \quad K = diag(K', K', \dots, K')$$

It is clear that the block-diagonal and upper bidiagonal matrix $\Delta = KDK^{-1} = diag(S(x_0, \dots, S(x_{n-1})$ has the desired form since:

$$S(x_i) = K'B(x_i)K^{'-1} = \begin{bmatrix} x_i & 1 & 0 & \dots & 0 \\ 0 & x_i & 1 & \dots & 0 \\ 0 & 0 & x_i & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 1 \\ 0 & 0 & \dots & \dots & x_i \end{bmatrix}$$

On the other hand, the transformation we perform over the Sylvester's structure consists simply in post multiplying it by the diagonal matrix K^{-1}. In doing so, one realizes that using the variable change, $Y = XK^{-1}$, the Sylvester's equation becomes:

$$Z^T Y - Y\Delta = e_{m-1} y^t K = e_{m-1} g^t$$

An interesting feature in this normalization step is that the blocks $S(x_i)$ in the block-diagonal matrix Δ are upper bidiagonal Toeplitz matrices each of which thereby commutes with any other upper triangular Toeplitz matrix. In the next transformation, we will make useful of this important property.
Consider n upper triangular and Toeplitz matrices $R_0, R_1, \dots, R_{n-1}$ each of order p, and set:

$$R = diag(R_0, R_1, \dots, R_{n-1}),$$

that is R is a block-diagonal matrix where each block-element is a $p \times p$ matrix and its block-main diagonal is formed by the introduced Toeplitz matrices R_i.

Clearly, if the structure is post-multiplied by R then remarking that $\Delta R = R\Delta$, we obtain:

$$Z^T YR - YR\Delta = e_{m-1}g^t R = e_{m-1}w^t$$

so that the variable change, $Y \to YR$, does not alter the overall required structure of the transformed Sylvester's equation. Therefore, it is possible in general to choose the matrices R_i in such a way that the coordinates w_{ip+k}, $k \neq 0$, of the row-vector

$$w^t = (w_0, w_1, \ldots, w_{np-1})$$

will be reduced to zero. It is readily seen that this operation fails if and only if there exists at least one coordinate of w of the form w_{ip} reduced to zero. Up to now, it is assumed that the transformations we may apply does not lead to this situation. It is necessary to note, by the way, that this normalization is reversible since the the Toeplitz matrices R_i can be reconstructed systematically.

Let us now proceed to determine explicitly the solution X (we will denote henceforth $W(x_0, \ldots, x_{n-1})$) of the Sylvester's equation:

$$Z^T X - X\Delta = -e_{m-1}y^t$$

$$y^t = (1,\ 0, \ldots, 0,\ 1,\ 0, \ldots, 0\ , \ldots, 1\ , 0, \ldots, 0)$$

in terms of the numbers x_i. Let a_{ih} denote the (i,h)-entry of the matrix X and write $h = jp + k$ such that $0 \leq j \leq n-1$ and $0 \leq k \leq p-1$. By controlling the elements of X in its structure, it is easily verified that:

$$\begin{cases} x_j a_{m-1,h} = 1 & (k = 0) \\ a_{m-1,h-1} + x_j a_{m-1,h} = 0 & (1 \leq k < p) \\ a_{m-i+1,h} - x_j a_{m-i,h} = 0 & (k = 0) \\ a_{m-i+1,h} - (a_{m-i,h-1} + x_j a_{m-i,h}) = 0 & (1 \leq k < p) \end{cases}$$

Clearly, these recurrence relations fulfilled by the elements of the matrix X provide the way of computing each $a_{m-i,h}$. Even more, they allow to show, using an inductive reasoning about (i,k), that:

$$a_{m-i,h} = (-1)^k \binom{k+i}{k+1} x_j^{-(k+i)}$$

Indeed, by considering the row $(a_{m-1,.})$ and the column $(a_{.,jp})$ of X, one realizes that $(a_{m-1,jp+k})_k$ and $(a_{m-i,jp})_i$ are geometrical progressions from which it is deduced that the formula is true for (i,k) such that $i = 1$ or $k = 0$. On the other hand, if we assume its trueness for $(i-1,k)$ and $(i,k-1)$ $(2 \leq i \leq m,\ 1 \leq k \leq p-1)$, and we set

$$b_{ih} = (-x_j)^k a_{m-i,h},$$

then after simple calculations, we can perform the operations below:

$$\begin{aligned} b_{i,h} &= (-x_j)^{k-1} a_{m-i,h-1} + (-1)^k x_j^{k-1} a_{m-i+1,h} \\ &= (-x_j)^{k-1}(-1)^{k-1}\binom{k-1+i}{k} x_j^{-(k-1+i)} + (-1)^k x_j^{k-1}(-1)^k \binom{k+i-1}{k+1} x_j^{-(k+i-1)} \\ &= \binom{k-1+i}{k} x_j^{-i} + \binom{k+i-1}{k+1} x_j^{-i} \\ &= \left(\binom{k+i-1}{k} + \binom{k+i-1}{k+1}\right) x_j^{-i} \\ &= \binom{k+i}{k+1} x_j^{-i} \end{aligned}$$

It follows directly that $a_{m-i,h} = (-1)^k x_j^{-k} b_{ih} = (-1)^k \binom{k+i}{k+1} x_j^{-(i+k)}$ and, as a consequence, the desired formula holds for all elements of the matrix X.

As an application to this result, it is possible to derive an expressive formula for the inverse of nonsingular $m \times m$ matrices which are solutions of the just established normalized Sylvester's structure. Indeed, if $W = W(x_0, \ldots, x_{n-1})$ is the solution of the normalized equation, then it is easy to check that:

$$W^{-1} Z^T - \Delta W^{-1} = W^{-1} e_{m-1} y^t W^{-1},$$

and by applying the anti-transpose operator T_2, we obtain:

$$Z^T W^{-T_2} - W^{-T_2} \Delta^{T_2} = W^{-T_2} \bar{y} e_1^t W^{-T_2}$$

$\bar{y}$ being the mirror vector of y. As a consequence, it is possible to assert that under the conditions where the normalization just presented does not fail, there exist an $m \times m$ lower triangular Toeplitz matrix L and n upper triangular Toeplitz matrices $R_0, R_1, \ldots, R_{n-1}$ each of order p, such that:

$$L^T W^{-T_2} R = W(x_{n-1}, x_{n-2}, \ldots, x_0)$$

where $R = diag(R_0, \ldots, R_{n-1})$.

6. Transformation to structured matrices. We consider here the aspect discussed in section 3 where it is shown that if the block-elements of a block-Vandermonde matrix V_b are symmetric matrices then $V_b V_b^T$ is a Hankel matrix (see [7]). In our situation however, this result might logically fail in general since the block Vandermonde matrix from which the confluent Vandermonde one is extracted has not the symmetry property: the block-elements are, rather, upper triangular matrices. To palliate this difficulty, we suggest the approach of looking forward matrices A such that $V_p A V_p^T$ is a Hankel or even a Hankel-like matrix. Let A be such a matrix, that is $V_p A V_p^T$ is Hankel:

$$Z^T V_p A V_p^T - V_p A V_p^T Z = a e_{m-1}^t - e_{m-1} a^t$$

Using the structures of V_p and V_p^T, it is readily verified under reasonable assumptions (V_p is a nonsingular square matrix) and after simple calculations that:

$$DA - AD^T = v u^t - u v^t$$

where $u = V_p^{-1} a + y, \quad v = V_p^{-1} e_{m-1}$. Therefore, it is natural to investigate the simple case where the right hand side is reduced to zero, i.e. to determine the matrices A for which $DA = AD^T$. If we set $B = KAK^{-1}$, where K is the diagonal matrix introduced in the preceding section, then the problem amounts to compute B verifying $B\Delta = \Delta B$. In this direction, we saw that the block-diagonal matrix Δ is formed by the upper bidiagonal Toeplitz matrices $S(x_i)$. Hence, it is not hard to show that the matrix

$$B = diag(R_0 J_p, R_1 J_p, \ldots, R_{n-1} J_p),$$

where the R_i are $p \times p$ upper triangular Toepliz matrices and J_p denotes the $p \times p$ reverse identity matrix, commutes with Δ. In the particular case where for all i, R_i is reduced to the $p \times p$ identity matrix, one finds that $V_p F V_p^T$ is a Hankel matrix where

$$F = K^{-1} diag(J_p, J_p, \ldots, J_p) K$$

Though this result permits to express V_p^{-1} or V_p^{-T} with the help of Hankel matrices deducing hence algorithms for calculating V_p^{-1} using those inverting a Hankel matrix, there is unfortunately a disadvantage making it somewhat meaningless. That is in the real case, the matrix $V_p F V_p^T$ is not positive definite, so that there is no way to compute the pseudo inverse or "the weighted one" of V^T using such a matrix. On the other hand, if we assume that $DA - AD^T = GH$ corresponding to the case where VAV^T is in general a Hankel-like matrix, it turns out that this situation yields more complicated symmetric matrices A.

As an alternative approach, we proceed now to construct matrices A for which VAV^T are Toeplitz-like, where it is assumed here that the $m \times np$ matrix $(m \leq np)$ V is real and satisfies the Sylvester's equation below:

$$Z^T V - V\Delta = e_{m-1} y^t$$

in such a way that VK is a confluent Vandermonde matrix, K being the diagonal matrix introduced in the previous section. On the other hand let construct the $np \times np$ matrix C defined as follows:

$$\bar{Z}^T C - C\Delta^{T_2} = g_{np-1} z^t$$

where $\bar{Z}$ stands for the $np \times np$ displacement matrix. It is clear that the vector z may be chosen such that CK is a nonsingular confluent Vandermonde matrix. Therefore if we set $W = C^{T_2}$ and $A = WW^T$ then one may claim that VAV^T is a Toeplitz-like matrix. To show this, let us write

$$X \equiv Y$$

for meaning that the difference $X - Y$ of the matrices X and Y is a summation of r outer products with r fixed. Thus using simple arguments, one finds that:

$$\begin{aligned} Z^T VAV^T Z &= Z^T VWW^T V^T Z \\ &\equiv V\Delta WW^T \Delta^T V^T \\ &\equiv VW\bar{Z}^T \bar{Z} W^T V^T \\ &\equiv VWW^T V^T \\ &= VAV^T \end{aligned}$$

which permits to conclude that VAV^T is indeed a Toeplitz-like matrix. In fact, it is possible to show that the product VW is a Toeplitz matrix so that its covariance matrix will be evidently Toeplitz -like. Indeed, one has just to post-multiply the structure

$$Z^T V - V\Delta = e_{m-1} y^t$$

of V by the introduced matrix W and then use the structure of C in order to find that of VW. It is not hard to check that:

$$\Delta W = W\bar{Z}^T - \bar{z}g_0^t$$

which can be substituted in the structure of V multiplied by W. In doing so, one finds that:

$$Z^T VW - VW\bar{Z}^T = e_{m-1}y^t W - V\bar{z}g_0^t$$

from which we conclude that our task is accomplished.

Besides the general interest of this result, it deserves to say that it is more significative than the just established one in the sense that A and VAV^T are symmetric positive definite matrices and the solution x of the linear system $VAV^T x = VWb$ solves the least squares minimization problem below:

$$min||W^T V^T x - b||_2$$

The importance of this observation lies in the fact that there are asymptotically fast methods in order to solve it. Indeed one may exploit the $O(np(\log np)\log n)$ fast method of [17] for computing the products VAV^T and VWb, and on the other hand it is possible to solve the symmetric positive definite and Toeplitz-like system $VAV^T x = VWb$ with $O(m\log^2 m)$ operations only appealing the rapid techniques developed in [8]. Hence, $O(np(\log np)\log n + m\log^2 m)$ arithmetic operations are sufficient for solving the just introduced least squares minimization problem.

7. Confluent Vandermonde least squares minimizations. In this section, we present another application of the Sylvester's structure of the $m \times np$, $(m \geq np)$, real confluent Vandermonde matrix V_p and show that it is possible to construct an $O((np)^2)$ algorithm for computing the Cholesky factorization of its covariance matrix $V_p^T V_p$. The idea is to establish a Sylvester's structure for the covariance matrix which can be achieved by computing $D^T V_p^T V_p D$ using the structures of V_p and V_p^T. We can write:

$$D^T V_p^T V_p D = (V_p^T Z - ye_{m-1}^t)(Z^T V_p - e_{m-1}y^t) = V_p^T ZZ^T V_p + yy^t$$

Observing that $ZZ^T = I_m - e_0 e_0^t$, and denoting by v^t the first row of V_p, it is readily verified that:

$$D^T V_p^T V_p D + vv^t = V_p^T V_p + yy^t$$

Let now $V_p^T V_p = R^T R$ be the Cholesky factorization of the covariance matrix $V_p^T V_p$, and let substitute it into this relation:

$$D^T R^T RD + vv^t = R^T R + yy^t$$

Therefore, we obtain what we can call a rank-2 modification of the Cholesky factorization with the difference that here we are concerned with the determination of the matrix R. It is therefore logical to appeal to the well known techniques used for and adapt them to our situation (see [9] for example). As it is customary in such a case, we will multiply the augmented matrices below:

$$\begin{bmatrix} y^t \\ R \end{bmatrix}, \quad \begin{bmatrix} v^t \\ RD \end{bmatrix}$$

by appropriate Givens rotations. The fundamental property of which we take advantage is that if a step in the method stands for a multiplication by a pair of Givens plane rotations, then each row of R (and RD) is modified only once. In what follows, we describe the first step where it is shown how the first row of R is determined. Let H_1 and H_2 be the following plane rotations:

$$H_1 = \begin{bmatrix} c & -s & 0 & \dots & 0 \\ s & c & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix}, \quad H_2 = \begin{bmatrix} \delta & -\gamma & 0 & \dots & 0 \\ \gamma & \delta & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix}$$

such that the first coordinates v'_0 and y'_0 of v' and y' respectively are reduced to zero, where:

$$H_1 \begin{bmatrix} v^t \\ RD \end{bmatrix} = \begin{bmatrix} v'^t \\ R'' \end{bmatrix}$$

and

$$H_2 \begin{bmatrix} y^t \\ R \end{bmatrix} = \begin{bmatrix} y'^t \\ R' \end{bmatrix}$$

According to the property stated above, one maintains that the first row of R'' and that of R' are identical, and moreover they remain unmodified during the subsequent Givens rotations. Taking into account this and the fact that $v'_0 = y'_0 = 0$, one may write the equations below:

$$\begin{cases} (a) \;\; cv_0 - sr_0x = 0 \\ (b) \;\; \delta y_0 - \gamma r_0 = 0 \\ (c) \;\; cr_0x + sv_0 = \delta r_0 + \gamma y_0 \end{cases}$$

where r^t denotes the first row of R, r_0 the first coordinate of r and $x = x_0$. It is readily seen that the last relation (relation (c)) can be also expressed as follows:

$$x^2 r_0^2 + v_0^2 = r_0^2 + y_0^2$$

from which one obtains the value of r_0 :

$$r_0 = \sqrt{\frac{y_0^2 - v_0^2}{x^2 - 1}},$$

as a consequence the numbers $c,\ s,\ \delta,\ \gamma$ can be calculated as follows:

$$d = \sqrt{r_0^2 + y_0^2},\ c = \frac{r_0 x}{d}, \quad s = \frac{v_0}{d}, \quad \delta = \frac{r_0}{d}, \quad \gamma = \frac{y_0}{d}$$

On the other hand, since the first row of R' and that of R'' are identical, it follows that:

$$cr^t D + sv^t = \delta r^t + \gamma y^t$$

and thereby:

$$r^t = (\gamma y^t - s v^t)(\delta I_{np} - cD)^{-1}$$

is the solution of a bidiagonal system. Of course, the row vectors v'^t and y'^t can be computed easily as a preparation for computing analogously the subsequent rows of R. We have:

$$v'^t = cv - sr^t D, \quad y'^t = \delta y - \gamma r^t$$

It is obvious that the upcoming steps for computing the rows of the Cholesky factor R may be realized identically since the shape of the augmented matrices

$$\begin{bmatrix} y^t \\ R \end{bmatrix}, \quad \begin{bmatrix} v^t \\ RD \end{bmatrix}$$

after each step may be rearranged in such a way to have the original one.

We have thus sketched a possible method for performing the Cholesky factorization of $V_p^T V_p$ using $O((np)^2)$ operations. However as it may be observed, it contains divisions and therefore care must be taken in its implementation. If it happens that a division by zero occurs for computing an unknown, we suggest as a solution the use of the classical Cholesky factorization algorithm in order to circumvent the problem. In this perspective, it is not hard to remark that if there are cosines c and δ corresponding to a same step, say the first one, and an index k such that $cx_k = \delta$ then the $(k+1)p^{th}$ coordinate $r_{(k+1)p-1}$ of r and only it is missed in the relation $cr^t D + sv^t = \delta r^t + \gamma y^t$ on which the computation of r is based. Since the rows of the factor R are determined by this method in an ascending order and each contains at most two coordinates in difficulty, it follows that the classical Cholesky factorization approach can be utilized successfully to compute $r_{(k+1)p-1}$ without altering the total computational time.

We end this discussion by noting that the guidelines of the Cholesky factorization method developed here can be used to compute the Cholesky factorization of the matrix $V_p^T S V_p$ where S is a symmetric positive definite Toeplitz matrix.

8. Generalization. In view of the formalization established in this paper, it is natural to consider what we may call the generalized confluent Vandermonde matrices defined as solutions of Sylvester's structures of the type:

$$Z^T X - X\Delta = GH$$

where in general the generator GH is a summation of α outer products, α being a (relatively small) positive integer, and Δ is the block-diagonal matrix introduced in section 4. Such a definition is precisely motivated by the fact that the linear operator:

$$\Phi : X \to \Phi(X) = Z^T X - X\Delta$$

is one-to-one, so that one is assured of the existence and uniqueness of the solution X of the equation $\Phi(X) = B$, no matter what is the right hand side matrix B. The problem that obviously arises in this context is to develop algorithms manipulating the generalized confluent Vandermonde matrices in a way similar to that available for the simple case where $\alpha = 1$. It is readily seen that the transformation to structured matrices shown in section 6 as well as the Cholesky factorization described in section 7 can be directly extended to this case. On the other hand, it follows directly from the linearity of the operator Φ and from the results of section 5, that if A is a generalized confluent Vandermonde matrix and the solution of the Sylvester's equation just established, then there are generally α lower triangular Toeplitz matrices $L_0, \dots, L_{\alpha-1}$ and α block-diagonal matrices $R^0, \dots, R^{\alpha-1}$ where each block is a $p \times p$ upper triangular Toeplitz matrix, such that:

$$A = \sum_{k=0}^{\alpha-1} L_k^T W(x_0, \dots, x_{n-1}) R^k$$

As a consequence, the complexity of realizing matrix vector multiplications with generalized confluent Vandermonde matrices is, up to the α multiplicative factor, the same as that of computing the confluent Vandermonde matrix vector multiplication. It is thus possible to construct fast algorithms by exploiting the fast methods developed in [17]. Combining this observation with the results of section 6, one may conclude that it is possible to construct asymptotically fast algorithms for computing the inverse of a nonsingular generalized confluent Vandermonde matrix, using the methods presented in [8].

References

[1] A.V. Aho, J.E.Hopcroft and J.D.Ullman. The design and analysis of computer algorithms *Addison Wesly, Reading, MA.*1973

[2] A. Ballester and V.Perreyra. On the construction of discrete approximations to linear differential expressions *Math. Comp.* 21(1967) pp.297-302

[3] G. Bitmead and B.D.O.Anderson. Asymptotically fast solution of Toeplitz and related systems *Lin. Alg. Appli.* 34(1980) pp.103-116

[4] A. Björk and T.Elfving. Algorithms for confluent Vandermonde systems *Numer. Math.* 21(1973) pp.130-137

[5] A. Björk and V.Perreyra. Solution of Vandermonde systems of equations *Math. Comp.* 24(1970) pp.893-903

[6] R. Brent, F.G.Gustavson and D.Y.Y.Yun. Fast solutions of Toeplitz systems of equations and computation of Padé approximants *J. Algorithms* 1(1980) pp.259-295

[7] J.F. Canny, E.Kaltofen and L.Yagati. Solving systems of non linear equations faster *in Proceeding ACM-SIGSAM 1989Internat. Symp. Symbolic Algebraic Comput. Assoc. Comput. Mach., NewYork* (1989) pp.34-42

[8] J. Chun and T. Kailath. Divide and conquer solutions of least squares problems for matrices with displacement structure *SIAM J. Matrix. Anal. Appli.* 12(1991) pp.128-145

[9] P.E.Gill, G.H.Golub,W.M.Murray, and Saunder's Methods for modifying matrix factorizations *Math.Comp.*28(1974) pp.505-535

[10] I. Gohberg and V.Olshevsky. Fast algorithms with preprocessing for matrix-vector multiplication problems *J. Complexity* 10(1994) pp.411-427

[11] G.H. Golub and C.F.VanLoan. Matrix Computations 3^{rd} *ed.The John Hopkin's Press, Baltimore* 1996

[12] G. Heinig and K.Rost. Algebraic Methods for Toeplitz-like matrices and operators *Akademic-Verlag, Berlin and Birkhaüser,* 1984

[13] N.J. Higham. Fast solution of Vandermonde-like systems involving orthogonal polynomials *IMA J. Numer. Anal.*8(1988) pp.473-486

[14] N.J. Higham. Stability analysis of algorithms for solving confluent Vandermonde-like systems *SIAM J. Matrix Anal. Appli.*11(1990) pp.23-41

[15] T. Kailath, S.Y.Kung and M.Morf. Displacement ranks of a matrix *Bull. Amer. Math. Soc* 1(1979) pp.769-773

[16] T. Kailath and A.H.Sayed. Displacement structure: theory and applications *SIAM review* 73(1995) pp.297-386

[17] H. Lu. Fast solution of confluent Vandermonde linear systems *SIAM J. Matrix Anal. Appli.* 15(1994) pp.1277-1289

[18] H. Lu. Solution of Vandermonde-like systems and confluent Vandermonde-like systems *SIAM J. Matrix Anal. Appli.* 17(1996) pp.127-138

[19] M. Morf Doubling algorithms for Toeplitz and related equations *in Proc. IEEE Internat. Conf. on Acoustics, speech and signal processing algorithms, Denver IEEE Press Piscataway, NJ,* (1980) pp.954-959

[20] W.P. Tang and G.H.Golub. The block decomposition of a Vandermonde matrix and its applications *BIT* 21(1981) pp.505-517