AN ALGEBRAIC APPROACH TO DISCRETE SHORT-TIME FOURIER TRANSFORM ANALYSIS AND SYNTHESIS

Z. Shpiro and D. Malah
Department of Electrical Engineering
Technion - Israel Institute of Technology
Haifa 32000, Israel

ABSTRACT

An algebraic representation of the discrete short-time Fourier transform (DSTFT) is presented for the case in which the analysis window length N equals the transform block size M. This representation allows the application of algebraic tools for determining an optimal synthesis system which minimizes the mean square error between a given modified DSTFT (which is not necessarily a valid DSTFT sequence) and the DSTFT of the synthesized signal. If no modification is applied, the result is a unity analysis-synthesis system for any given time update R of the sliding analysis window (provided that $R \leq M$). It is shown that the optimal synthesis system can be implemented by the well known weighted overlap-add (WOLA) method using an optimal synthesis window. The algebraic approach enables the extension of some recent results and the relaxation of a constraint on the analysis window. The proposed approach is found also to have a potential for solving the synthesis problem for the more general case of $N>M$.

I. INTRODUCTION

In recent years the discrete short-time Fourier transform (DSTFT) has proved to be a useful tool for the analysis, modification, and synthesis of nonstationary signals, such as speech [1-7]. Portnoff [4] has extended the earlier mathematical models and introduced the explicit use of a synthesis window. Crochiere [5] developed the weighted overlap-add (WOLA) method for efficient reconstruction of a signal from a given DSTFT sequence by weighting each inverse-transformed block by the synthesis window and overlap-adding the weighted blocks. Typically, the analysis window is selected to satisfy specifications given on the performance of the analysis filter-bank. However, the problem of finding the appropriate synthesis window is yet not fully solved. A necessary and sufficient condition for exact reconstruction (if no modification of the input signal DSTFT is performed) was derived by Portnoff [4] but did not lead to an explicit design method. The first step in this direction was taken by Griffin and Lim [8], who considered the synthesis of a signal from a modified STFT (which differs from the DSTFT considered here in that the frequency variable is continuous), under the criterion of minimizing the mean square error between the given modified STFT (which may not be a valid STFT) and the STFT of the synthesized signal. If no modification is performed a unity system is obtained. However, not only is their result proved for the STFT and not the DSTFT, but it is also derived under the assumption that the analysis window $h(n)$ is not zero in the range of its definition, i.e. $n=0,1,\ldots,N-1$, where N is the analysis window length.

Since in any computer implementation of an analysis-synthesis system the frequency variable must also be discretized, we attempted to extend the results obtained in [8] also for the DSTFT. Using an algebraic representation of the DSTFT analysis and synthesis operations, we show in this paper that if the number of discrete frequencies M equal the window length N, the results obtained in [8] are also valid here. Furthermore, we show that the result in [8] can be put in the form of WOLA synthesis, and that the assumption that the analysis window has no zero values in its range of definition can be greatly relaxed. The algebraic approach used in this paper can be applied also when the analysis window length $N>M$. Such windows are needed if the amount of overlap of adjacent analysis filters in the frequency domain is to be reduced [6,7]. In this paper we consider only the case $N=M$, since if $N>M$ the problem is more difficult. Preliminary results obtained for $N>M$ are reported in [9].

Similar to [8], we also consider here the situation of synthesizing a signal from a given modified DSTFT (MDSTFT). The DSTFT of a signal $x(n)$ is defined by:

$$X(sR,k\Delta\Omega) \triangleq \sum_{m=-\infty}^{\infty} x(m)h(sR-m)exp(-jk\Delta\Omega m)$$

The optimal synthesis system is defined to be the one minimizes the mean square error ε^2:

$$\varepsilon^2 \triangleq \frac{1}{N}\sum_{s=-\infty}^{\infty}\sum_{k=0}^{N-1}|\hat{X}(sR,k\Delta\Omega)-Y(sR,k\Delta\Omega)|^2 \qquad (1)$$

where $\Delta\Omega=2\pi/N$, R is the time-update (in samples) of the sliding analysis window, $\hat{X}(sR,K\Delta\Omega)$ is the DSTFT of the reconstructed signal $\hat{x}(n)$, and $Y(sR,k\Delta\Omega)$ is the given MDSTFT sequence.

The outline of the paper is as follows: In the next section the algebraic representation of DSTFT analysis and WOLA synthesis are formulated. In Section III the optimal synthesis system is found using algebraic tools, and in Section IV we summarize the results and present conclusions.

II. Algebraic Representation of DSTFT Analysis and Synthesis

The algebraic representation to be presented in this section is directly based on the implementation scheme of DSTFT analysis and weighted overlap-add (WOLA) synthesis developed in [5]. To conserve space, we describe this implementation scheme only briefly. According to this scheme an input segment is weighted by the analysis window of length N and is then transformed using a DFT of length N (via the FFT). The short-time transform obtained this way is referenced to the beginning of the sliding analysis window (sliding time reference) and hence must be aligned with the fixed time origin. This can be accomplished either in the time domain - before the transformation - by circularly rotating the windowed data, or in the frequency domain by applying a linear phase shift. The fixed time reference DSTFT can now be modified (if necessary). The synthesis (or reconstruction) is performed by first returning to the sliding time reference and then applying an inverse DFT (IDFT). The resulting signal block is weighted by the synthesis

CH1945-5/84/0000-0015 $1.00 © 1984 IEEE

window and overlapped-added to the output buffer. The processing is continued block-by-block with a time-update of R samples (typically $1<R<N$).

The above block-by-block implementation can be represented in a matrix form as follows. Let $\underline{x}$ denote the input signal vector, having as its elements the samples of the input sequence $x(n)$. Let $\underline{b}$ denote the fixed time reference DSTFT vector which results by concatenating transform blocks. Then, the relation between $\underline{b}$ and $\underline{x}$ can be written in matrix form as

$$C\underline{x} = \underline{b} \tag{2}$$

where the matrix C performs the earlier described processing steps on the input vector. If $R<N$, the matrix C is not square but rectangular, with dimensions $m\times l$, with $m>l$. The matrix C can be decomposed into three factor matrices in two ways, according to the time reference alignment used, i.e.,

$$C = FPA \tag{3}$$

if the time reference alignment is done in the time domain (by the matrix P), or

$$C = QFA \tag{4}$$

if this is done in the frequency domain (by the matrix Q), by introducing the needed linear phase shift. The matrix A which appears in (3) and (4) is a rectangular matrix which performs the windowing of consecutive input signal segments (each of length N and overlapping the previous input segment by $N-R$ points). The matrix F is a square matrix which implements the DFT of the consecutive blocks, and hence is a block-diagonal matrix. Each block W in F is of dimension $N\times N$ and is the usual DFT matrix with elements $w_{i,k}$ given by

$$w_{i,k} = exp(-j\frac{2\pi}{N}ik) \quad , \quad \begin{matrix} i=0,1,\ldots,N-1 \\ k=0,1,\ldots,N-1 \end{matrix} \tag{5}$$

Returning to the time reference alignment matrices P and Q above; both are square matrices. The matrix P is a block-diagonal permutation matrix which performs the circular rotation of each windowed input segment. Denoting the $N\times N$ s-th block by $[P^s]$, its elements are given by

$$p^s_{i,k} = \begin{cases} 1 & \text{if } ((k-i))_N = ((-sR-N/2))_N \\ 0 & \text{otherwise} \end{cases} \quad \begin{matrix} i=0,1,\ldots,N-1 \\ k=0,1,\ldots,N-1 \end{matrix} \tag{6}$$

where $(( ))_N$ denotes modulo N operation. The matrix Q is a diagonal matrix which provides the linear phase shift which is equivalent to the circular rotation provided by P. The diagonal of Q is composed of concatenated vectors $\underline{Q}^s$ of length N each, where the k-th element of the s-th vector is given by

$$q^s_k = (-1)^k \exp(-j\frac{2\pi}{N}sRk) \quad , \quad k=0,1,\ldots,N-1 \tag{7}$$

Fig. 1 illustrates the algebraic representation in (2) for an input vector $\underline{x}$ of length $l=8$, an analysis window $h(n)$ of length $N=4$ $(n=0,1,2,3)$, and a time update step-size of $R=2$ samples. The dimension of A is here $[(\frac{l-N}{R}+1)N]\times[l]=12\times8$, and each of the square matrices P,F, and Q are 12×12. Note that as the input vector gets longer the matrix C expands.

Similar to the above description of the analysis process, the synthesis process, according to the WOLA approach, takes the form

$$D\underline{b} = \underline{\hat{x}} \tag{8}$$

where $\underline{b}$ is a given DSTFT vector (possibly a modified version of the DSTFT of an input signal), $\underline{\hat{x}}$ is the synthesized signal vector, and D is a matrix which performs the WOLA synthesis process described earlier. If $R<N$, D is a rectangular matrix of dimension $l\times m$, $m>l$. As in the analysis stage, D can be factored into a product of three matrices, either as

$$D = S\hat{P}\hat{F} \tag{9}$$

or as

$$D = S\hat{F}\hat{Q} \tag{10}$$

Equation (9) is the counterpart of (3), and (10) is the counterpart of (4). The matrices $\hat{P},\hat{F},\hat{Q}$ are square matrices which perform the inverse operations of those performed by P,F,Q, respectively, in the analysis stage. Hence, $\hat{F}$ is a block diagonal matrix with each $N\times N$ block $[\hat{W}]$ performing the IDFT. $\hat{P}$ is a block diagonal matrix which performs the reverse circular rotation performed by P, and $\hat{Q}$ cancels the linear phase shift introduced by Q. Thus, we have

$$\hat{P}P = I \tag{11}$$

$$\hat{Q}Q = I \tag{12}$$

and

$$\hat{F}F = I \tag{13}$$

The matrix S is a rectangular matrix (if $R<N$) which performs the multiplication of consecutive signal blocks by the synthesis window $f(n)$, and the overlap-add operation. Fig. 2 illustrates the synthesis stage by the above matrix operations, for the example described in Fig. 1. It is noted from the example that S is identical to the transpose of A (i.e. A^T), if the samples of the analysis window $h(n)$ are replaced by the samples of the synthesis window $f(n)$.

III. Optimal Synthesis

As discussed earlier, the analysis window shape and its parameters (namely, N and R) are typically dictated by the application at hand. The problem then is to determine the synthesis system which minimizes the mean square error ε^2 defined in (1). With this error criterion, if the modified DSTFT sequence is a valid DSTFT, the optimal synthesis system should give $\varepsilon^2=0$. A particular case of this situation is when no DSTFT modification is applied and therefore the optimal analysis-synthesis system is a unity system.

Since the algebraic representation of the analysis stage takes the form of a set of linear equations, i.e., $C\underline{x}=\underline{b}$, the synthesis problem is to solve for $\underline{x}$, given C and $\underline{b}$. The general solution we seek is of the form $\underline{\hat{x}}=\hat{D}\underline{b}$. The problem is how to find $\hat{D}$ when C is rectangular and of very large dimensions (tending to infinity), such that ε^2 in (1) be minimum. Observing that $\hat{X}$ in (1) is given by $C\underline{\hat{x}}$ we obtain that minimizing ε^2 is equivalent to minimizing $\hat{\varepsilon}^2$ defined below:

$$\hat{\varepsilon}^2 \triangleq ||C\underline{\hat{x}} - \underline{b}||^2 \tag{14}$$

where $||\underline{u}||$ denotes the norm of the vector $\underline{u}$. It should be noted that although the algebraic representation of the WOLA synthesis described in the previous section has the form $\underline{\hat{x}}=D\underline{b}$, (see (8)), the operation of the optimal synthesis matrix $\hat{D}$ is not necessarily implementable by the WOLA approach, unless $\hat{D}$ can be factored into a product of three matrices as defined in (9) and (10). It will be seen in the sequel that for the problem under discussion (i.e., the window length N is equal to the transform block size), the solution to be forwarded for $\hat{D}$ can be put in a WOLA form with an appropriate (optimal) synthesis window.

The solution for $\hat{D}$ depends on the form of the analysis matrix C which is determined by the analysis stage parameters N and R and the shape of analysis window. Let us consider first the simplest situation of $R=N$. In this case the analysis is done block by block, without overlap. C is now a square matrix, A a square-diagonal matrix and P the identity matrix $P=I$ (see (6)). The diagonal of A is periodic with period N and its elements are

given by the samples of the analysis window. Thus, assuming that C is invertible, which is the case if the analysis window $h(n)$ satisfies $h(n)\neq 0$, $n=0,1,...,N-1$, we get $\hat{D}=C^{-1}$ and hence, using (3) and $P=I$, we obtain

$$\underline{\hat{x}} = C^{-1}\underline{b} = (FPA)^{-1}\underline{b} = A^{-1}F^{-1}\underline{b} \qquad (15)$$

Comparing this solution with (9) we have $S=A^{-1}$ (which is also diagonal), $\hat{P}=P=I$, and $\hat{F}=F^{-1}$ is indeed the IDFT matrix. The optimal synthesis takes therefore the form of WOLA synthesis with a synthesis window $f(n)$ given by $f(n)=1/h(n)$, $n=0,1,...,N-1$.

We turn now to the more general case for which $R<N$. Since C has now more rows than columns ($m>l$), there are infinitely many possible solutions. If $\underline{b}$ represents a valid DSTFT sequence (i.e. there exists a vector $\underline{y}$ such that $C\underline{y}=\underline{b}$), any solution which gives $\hat{D}C=I$ ($l\times l$) is acceptable. However, if $\underline{b}$ does not represent a valid DSTFT sequence we need to find a solution which minimizes $\hat{\varepsilon}^2$ in (14). Such a solution is given by the generalized inverse [10] of C, denoted by $C^{\dagger}$ Thus, the optimal synthesis is described by

$$\underline{\hat{x}} = \hat{D}\underline{b} \qquad (16)$$

with

$$\hat{D} = C^{\dagger} = (C^{*}C)^{-1}C^{*} \qquad (17)$$

where C^{*} denotes the conjugate transpose of C (C is defined over the complex field and is assumed here to have a rank which is equal to the number of its columns). It is noted that $C^{\dagger}C=I$ and hence this solution is also acceptable if $\underline{b}$ represents a valid DSTFT as discussed above. It also coincides with (15) if $R=N$.

Our work is yet incomplete since we have to find a way for computing $C^{\dagger}$, which has dimensions which tend to infinity, and relate the solution to the WOLA synthesis method (or else give a physical interpretation to the synthesis process). Let us examine first the process of computing $C^{\dagger}$: Using (3) and (17) we get

$$C^{\dagger} = (A^{T}P^{T}F^{*}FPA)^{-1}A^{T}P^{T}F^{*} \qquad (18)$$

But, $P^{T}=P^{-1}$; $F^{*}F=NI$ (or $F^{*}=NF^{-1}$), and hence

$$C^{\dagger} = (A^{T}A)^{-1}A^{T}P^{T}F^{-1} \qquad (19)$$

and thus

$$\underline{\hat{x}} = C^{\dagger}\underline{b} = [(A^{T}A)^{-1}A^{T}]P^{T}F^{-1}\underline{b} \qquad (20)$$

The interpretation of (20) is that the given DSTFT vector $\underline{b}$ is first inverse transformed by F^{-1} (IDFT), then it is circularly rotated to return to the sliding time reference, and finally multiplied by $\hat{S}$, where

$$\hat{S} \triangleq (A^{T}A)^{-1}A^{T} \qquad (21)$$

The first two operations are identical to those described in the previous section for the WOLA synthesis (see (9)), so that we can identify P^{T} with $\hat{P}$ and F^{-1} with $\hat{F}$ and hence the intermediate signal vector $\underline{v}$ defined below is the same for both synthesis methods.

$$\underline{v} \triangleq P^{T}F^{-1}\underline{b} = \hat{P}\hat{F}\underline{b} \qquad (22)$$

We describe now two ways for completing the synthesis process. The first approach is to initially generate a signal vector $\underline{x}'$ defined by

$$\underline{x}' \triangleq A^{T}\underline{v} \qquad (23)$$

This is equivalent to a WOLA operation on $\underline{v}$ with a synthesis window which is identical to the given analysis window (recall that S in (9) has the form of A^{T}). The output signal vector $\underline{\hat{x}}$ is now found by weighting $\underline{x}'$ with the matrix $(A^{T}A)^{-1}$. It turns out that because of the special form of A (as illustrated in Fig. 1), $A^{T}A$ is a diagonal matrix with a diagonal which is periodic with period R, except for the first and last $N-R$ terms which relate to the transient phenomena for any given finite input signal vector $\underline{x}$. Since the dimensions of A are very large (tend to infinity) this transient phenomena is of no concern. The i-th term in each period on the diagonal is given by

$$\lambda_i = \sum_{m=-\infty}^{\infty} h^2(i-mR) \quad i=0,1,...,R-1 \qquad (24)$$

Hence, the matrix $(A^{T}A)^{-1}$, needed to compute $\underline{\hat{x}}$ from $\underline{x}'$, is also diagonal with the i-th element in each period given by $1/\lambda_i$. The final result is that the output sequence $\underline{\hat{x}}(n)$ is given by

$$\hat{x}(n) = \frac{1}{\sum_{m=-\infty}^{\infty} h^2(n-mR)}\, x'(n) \qquad (25)$$

The R values of λ_i in (24) (which also appear in the denominator of (25)) can be precomputed from the known values of the given analysis window $h(n)$ (which has only a finite length of N), and hence the reconstruction of $\hat{x}(n)$ from $x'(n)$ can be done on a sample by sample basis.

The second approach for completing the synthesis is to apply $\hat{S}$ of (21) directly to $\underline{v}$. Because A^{T} has the form of a synthesis matrix for the WOLA method and $(A^{T}A)^{-1}$ is diagonal, $\hat{S}$ has also the form of a synthesis matrix for the WOLA method, with the optimal synthesis window given by

$$f(n) = \frac{h(n)}{\sum_{m=-\infty}^{\infty} h^2(n-mR)} \qquad (26)$$

The WOLA synthesis is described in scalar form by [5,8]:

$$\hat{x}(n) = \sum_{s=-\infty}^{\infty} f(n-sR)\, y(n-sR,s) \qquad (27)$$

where $y(n-sR,s)$ is given from the elements of $\underline{v}$ in (22) by

$$y(n-sR,s) = v[\,((n-sR))_N + sN] \qquad (28)$$

Substituting (26) in (27) and introducing a change of variables: $r=s+m$, we obtain the alternative form

$$\hat{x}(n) = [\sum_{s=-\infty}^{\infty} h(n-sR)y(n-sR,s)] / \sum_{r=-\infty}^{\infty} h^2(n-sR) \qquad (29)$$

This is the result (with some change of notation) obtained by Griffin and Lim in [8].

Finally, we would like to discuss the assumption made earlier that the rank of the $m\times l$ matrix C is equal to l. From (3), C is given by the product of P,F and A. But, the matrices P and F are square ($m\times m$) nonsingular matrices, and hence the rank of C equals to the rank of the rectangular ($m\times l$) matrix A. Because A has only one single term in each row (a sample of the analysis window), a necessary and sufficient condition for A to have rank l is that none of its column vectors be identically zero. Because of the special form of A (see illustration in Fig. 1), each column is a polyphase filter [7] of the given analysis window (which is the impulse response of the lowpass prototype filter of the analysis filterbank). Hence, the condition assumed in [8] that the analysis window must be nonzero in the range $n=0,1,...,N-1$ can actually be relaxed and replaced by the condition that no polyphase filter of the analysis window be identically zero.

IV. Conclusions

An algebraic representation of DSTFT analysis and synthesis was presented in this paper for the particular case in which the analysis window length N equals the transform block size M. Based on this representation and with the help of algebraic tools an optimal synthesis system under a minimum mean error criterion was derived. If no modification is applied to the input signal DSTFT the resulting analysis-synthesis system is shown

to be a unity system for any given value of the time update step-size R, $R \leq M$, of the sliding analysis window. The resulting synthesis formulation is found to coincide with the result forwarded by Griffin and Lim in [8] which considered the simpler case of a continuous frequency variable. We show that the assumed constraint in [8] that the window has no zero value in its range of definition can be greatly relaxed, and it is sufficient that no polyphase filter of the given analysis window is identically zero. The relation of the optimal synthesis system to the well known WOLA synthesis method developed by Crochiere is studied and it is shown that the optimal synthesis system can be formulated also as a WOLA system with an appropriate optimal synthesis window. This shows that the synthesis method obtained in [8] is also a WOLA technique a fact which was not noticed by the authors of [8].

The use of the algebraic approach is not limited to the case considered in this paper, and the synthesis window which assures a unity analysis-synthesis system (when no modification is applied to the input DSTFT) was already found in [9] for the case in which the analysis window is longer than the transform block size (i.e. $N>M$). The problem of finding the optimal synthesis system under a minimum mean square error criterion when $N>M$ is now under study.

References

[1] J.B. Allen and L.R. Rabiner, "A Unified Approach to Short-Time Fourier Analysis and Synthesis", Proc. IEEE, Vol. 65, pp. 1558-1564, Nov. 1977.

[2] L.R. Rabiner and R.W. Schafer, Digital Processing of Speech Signals, Prentice-Hall, NJ, 1978.

[3] J.M. Tribolet and R.E. Crochiere, "Frequency Domain Coding of Speech", IEEE Trans. Acoust., Speech, Signal Processing, Vol. ASSP-27, pp. 512-530, Oct. 1979.

[4] M.R. Portnoff, "Time-Frequency Representation of Digital Signals and Systems Based on Short-Time Fourier Analysis", IEEE Trans. Acoust., Speech, Signal Processing, Vol. ASSP-28, pp. 55-69, Feb. 1980.

[5] R.E. Crochiere, "A Weighted Overlap-Add Method of Short-Time Fourier Analysis/Synthesis", IEEE Trans. Acoust., Speech, Signal Processing, Vol. ASSP-28, pp. 99-102, Feb. 1980.

[6] D. Malah and J.L. Flanagan, "Frequency Scaling of Speech Signals by Transform Techniques", Bell System Technical Journal, Vol. 60, No. 9, pp. 2107-2156, Nov. 1981.

[7] R.E. Crochiere and L.R. Rabiner, Multirate Digital Signal Processing, Prentice-Hall, NJ, 1983.

[8] D.W. Griffin and J.S. Lim, "Signal Estimation from Modified Short-Time Fourier Transform", Proc. 1983 IEEE Int. Conf. Acoust., Speech, Signal Processing, (ICASSP-83), pp. 804-807, April 1980.

[9] Z. Shpiro, Analog Speech Scrambling by Means of Discrete Short-Time Fourier Transform, M.Sc. Thesis (in Hebrew), Technion - Israel Institute of Technology, Haifa, Israel, Nov. 1983.

[10] A. Ben-Israel and T.N.E. Greville, Generalized Inverse Theory and Applications, John Wiley & Sons, 1974.

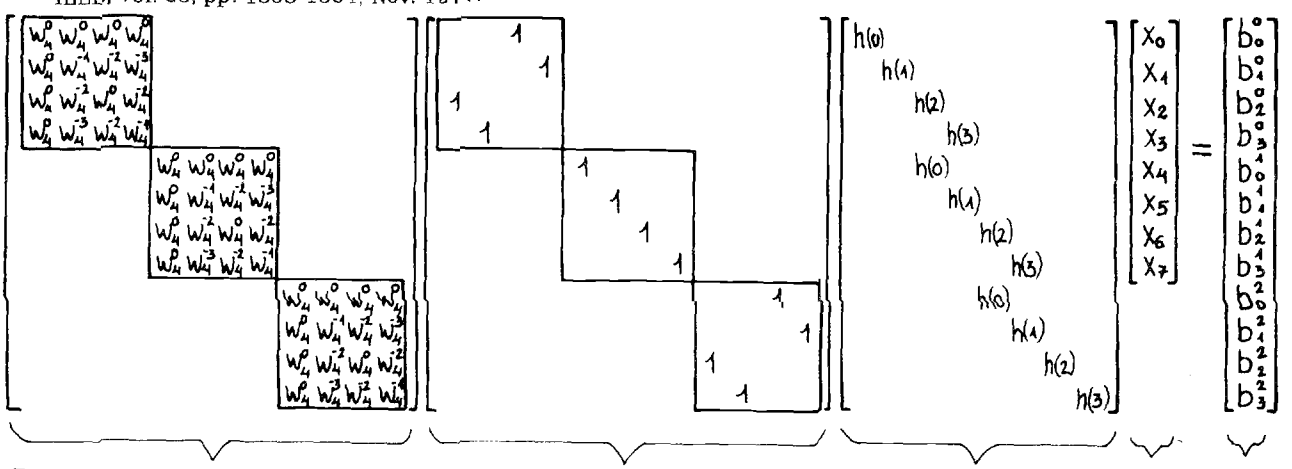

Fig. 1:

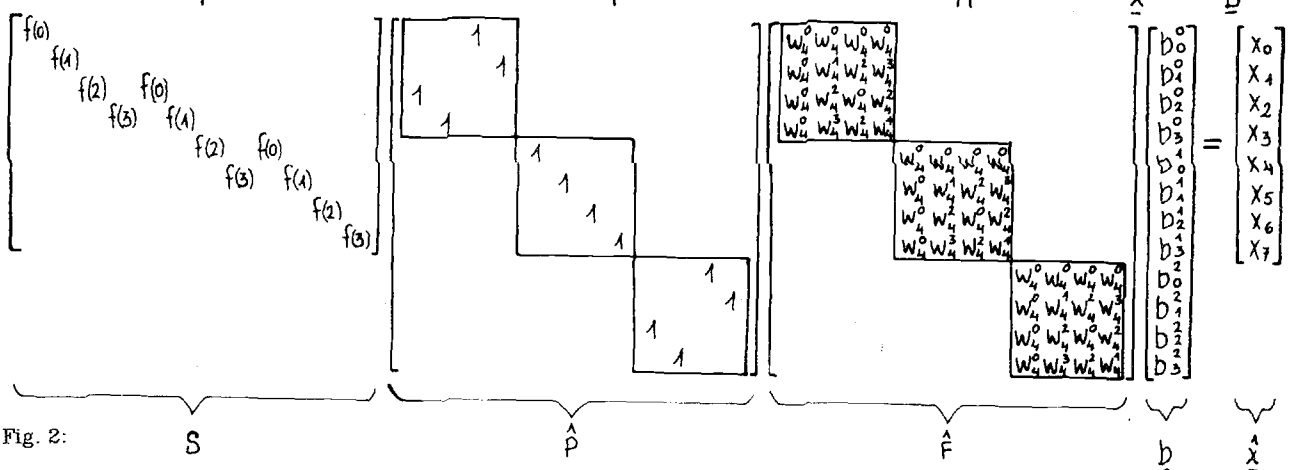

Fig. 2: