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 $R=M$). It is shown that the optimal synthesis system can be implemented by the well known weighted overlap-add (WOA) 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 (WOA) 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. $n=0,1,...,N-1$, 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 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:

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

The optimal synthesis system is defined to be the one minimizing the mean square error $e^2$:

$$e^2 = \frac{1}{N} \sum_{n=-\infty}^{\infty} \sum_{k=0}^{M-1} |X(sR,k\Delta) - Y(sR,k\Delta)|^2$$

where $\Delta = \pi/N$, $R$ is the time-update (in samples) of the sliding analysis window, $X(sR,k\Delta)$ is the DSTFT of the reconstructed signal $x(n)$, and $Y(sR,k\Delta)$ 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 (WOA) 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
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 \( z \) denote the input signal vector, having as its elements the samples of the input sequence \( x(n) \). Let \( b \) denote the fixed time reference DSTFT vector which results by concatenating transform blocks. Then, the relation between \( b \) and \( z \) can be written in matrix form as:

\[
C = F P A
\]

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.,

\[
C = Q F A
\]

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

\[
C = Q P A
\]

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 the 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 \( \tilde{F} \) in \( F \) is of dimension \( N \times N \) and is the usual DFT matrix, with elements \( w_{k} \) given by

\[
w_{k} = \exp(-j \frac{2 \pi}{N} k l), \quad k = 0, 1, \ldots, N - 1
\]

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

\[
P_{s} = \begin{cases} 1, & \text{if } (k-l)_{N} = ((k-R-N/2))_{N}, \quad l = 0, 1, \ldots, N-1 \\ 0, & \text{otherwise} \end{cases}
\]

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

\[
Q^{k}_{s} = (-1)^{s} \exp(-j \frac{2 \pi}{N} s \Delta k), \quad k = 0, 1, \ldots, N - 1
\]

Fig. 1 illustrates the algebraic representation in (2) for an input vector \( z \) 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 \( [(l-N+1)N]x[1]=128x6 \), and each of the square matrices \( P, F, \) and \( Q \) are \( 12 \times 12 \). Note that as the input vector gets longer, the matrix \( A \) expands.

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

\[
D = \tilde{F} P
\]

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

\[
D = \tilde{F} P
\]

or as

\[
D = \tilde{F} Q
\]

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

\[
\tilde{P} = I
\]

(11)

\[
\tilde{Q} = I
\]

(12)

and

\[
\tilde{F} = I
\]

(13)

The matrix \( S \) is a rectangular matrix (if \( R < N \)) which performs the multiplication of consecutive signal blocks by the synthesis window \( h(n) \), 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 \( e^{2} \) defined in (1). With this error criterion, if the modified DSTFT sequence is a valid DSTFT, the optimal synthesis system should give \( e^{2} = 0 \). 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., \( C \tilde{z} = b \), the synthesis problem is to solve for \( z \), given \( C \) and \( b \). The general solution we seek is of the form \( z = D \tilde{z} \). The problem is then to find \( D \) when \( C \) is rectangular and of very large dimensions (tending to infinity), such that \( e^{2} \) in (1) is zero. Observing that \( X \) in (1) is given by \( \tilde{z} \), we obtain that minimizing \( e^{2} \) is equivalent to minimizing \( e^{2} \) defined below:

\[
e^{2} = \frac{1}{N} | X - \tilde{z} |^{2}
\]

(14)

where \( | \tilde{u} | \) denotes the norm of the vector \( \tilde{u} \). It should be noted that although the algebraic representation of the WOLA synthesis described in the previous section has the form \( z = D \tilde{z} \) (see (8)), the operation of the optimal synthesis matrix \( D \) is not necessarily implementable by the WOLA approach, unless \( D \) 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 \( D \) can be put in a WOLA form with an appropriate (optimal) synthesis window.

The solution for \( \tilde{D} \) 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) ≠ 0, n = 0, 1, ..., N−1, we get D = C⁻¹ and hence, using (3) and P = I, we obtain

$$\tilde{x} = Cb = (FP)A^{-1}b = A^{-1}F^t b$$

(15)

Comparing this solution with (9) we have S = A⁻¹ (which is also diagonal), P = F = I, and hence 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 > n), there are infinitely many possible solutions. If $b$ represents a valid DSTFT sequence (i.e. there exists a vector $y$ such that $Cy = b$), any solution which gives $DC = f(XX^t)$ is acceptable. However, if $b$ does not represent a valid DSTFT sequence, we need to find a solution which minimizes $\tilde{x}^2$ in (14). Such a solution is given by the generalized inverse [10] of $C$, denoted by $C^t$.

Thus, the optimal synthesis is described by

$$\tilde{x} = \tilde{D}b$$

(16)

with

$$\tilde{D} = C^t = (C^t C)^{-1} C^t$$

(17)

where $C^t$ 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 $C^t C = f(XX^t)$ is acceptable. However, if $b$ does not represent a valid DSTFT sequence, we need to find a solution which minimizes $\tilde{x}^2$ in (14). Such a solution is given by the generalized inverse [10] of $C$, denoted by $C^t$.

Thus, the optimal synthesis is described by

$$\tilde{x} = \tilde{D}b$$

(16)

with

$$\tilde{D} = C^t = (C^t C)^{-1} C^t$$

(17)

where $C^t$ 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 $C^t C = f(XX^t)$ is acceptable. However, if $b$ does not represent a valid DSTFT sequence, we need to find a solution which minimizes $\tilde{x}^2$ in (14). Such a solution is given by the generalized inverse [10] of $C$, denoted by $C^t$.

Thus, the optimal synthesis is described by

$$\tilde{x} = \tilde{D}b$$

(16)

with

$$\tilde{D} = C^t = (C^t C)^{-1} C^t$$

(17)

where $C^t$ 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 $C^t C = f(XX^t)$ is acceptable. However, if $b$ does not represent a valid DSTFT sequence, we need to find a solution which minimizes $\tilde{x}^2$ in (14). Such a solution is given by the generalized inverse [10] of $C$, denoted by $C^t$.

Thus, the optimal synthesis is described by

$$\tilde{x} = \tilde{D}b$$

(16)

with

$$\tilde{D} = C^t = (C^t C)^{-1} C^t$$

(17)

where $C^t$ 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 $C^t C = f(XX^t)$ is acceptable. However, if $b$ does not represent a valid DSTFT sequence, we need to find a solution which minimizes $\tilde{x}^2$ in (14). Such a solution is given by the generalized inverse [10] of $C$, denoted by $C^t$.

Thus, the optimal synthesis is described by

$$\tilde{x} = \tilde{D}b$$

(16)

with

$$\tilde{D} = C^t = (C^t C)^{-1} C^t$$

(17)

where $C^t$ 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 $C^t C = f(XX^t)$ is acceptable. However, if $b$ does not represent a valid DSTFT sequence, we need to find a solution which minimizes $\tilde{x}^2$ in (14). Such a solution is given by the generalized inverse [10] of $C$, denoted by $C^t$.

Thus, the optimal synthesis is described by

$$\tilde{x} = \tilde{D}b$$

(16)

with

$$\tilde{D} = C^t = (C^t C)^{-1} C^t$$

(17)

where $C^t$ 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 $C^t C = f(XX^t)$ is acceptable. However, if $b$ does not represent a valid DSTFT sequence, we need to find a solution which minimizes $\tilde{x}^2$ in (14). Such a solution is given by the generalized inverse [10] of $C$, denoted by $C^t$.
to be a unity system for any given value of the time update step-size $R$, $R \geq M$, 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


