周波数間引き短時間フーリエ変換の提案
Proposal of Frequency-Undersampled Short-Time Fourier Transform

北原 大地
立命館大学 情報理工学部

Daichi KITAHARA
College of Information Science and Engineering, Ritsumeikan University

あらまし　短時間フーリエ変換 (Short-Time Fourier Transform: STFT) を用いて信号を解析する際，通常は隣接する時間フレームを半分以上重ねながら，フレーム長と同数の周波数成分を計算する．変換後のスペクトログラム行列の成分数は元の信号の成分数の 2 倍以上となるため，STFT を信号圧縮に用いることは滅多にない．また，スペクトログラムを所望の値に設定しても，STFT の値域に属さない限り，逆変換時に設定値からずれてしまう．本研究では，スペクトログラム行列の成分数の過剰な増加を防ぐため，フレーム長の半分の数の周波数成分のみを求める，周波数間引き STFT を提案する．性質の異なる 2 種類の逆変換を提示した後に，数値実験で提案手法の正当性を確認する．

1　はじめに

時間周波数解析とは，信号に含まれる周波数成分の時間変化を調べることである．短時間フーリエ変換 (Short-Time Fourier Transform: STFT) は，音信号に対して最もよく利用される時間周波数解析手法である [1]–[5]．信号に STFT を適用した結果は「スペクトログラム」と呼ばれ，行列の形式で表現されることが多い．信号の解析や特徴抽出のためにスペクトログラムを用いる以外にも，スペクトログラム自体を加工することで所望の時間信号を生成する場合も多い．本研究では，後者の利用法を強く意識する．

工学分野で STFT を用いる際には，対象が離散時間信号であり，更に窓関数がコンパクトな台を持つ場合がほとんどである．この場合には，離散時間窓関数の台の長さ (窓長) が，そのまま各時間フレームの長さ (フレーム長) となる．通常は，各フレームの信号からフレーム長と同数の周波数成分を高速フーリエ変換 (Fast Fourier Transform: FFT) により求める．本稿では，これを離散 STFT と呼ぶ．FFT の直前で各フレームの信号の末尾に 0 を埋め込めば，フレーム長よりも多くの周波数成分を計算することも可能である．本稿では文献 [6] での名称に倣い，これを周波数冗長化 STFT (Frequency-Oversampled STFT: FOSTFT) と呼ぶ．離散 STFT や FOSTFT の (疑似) 逆変換は，いわゆる双対窓の窓長もフレーム長と等しくなるため，順変換と同様の計算で実現可能であることがよく知られている [7]．

離散 STFT や FOSTFT を用いる際，人間の聴覚が音信号のブロック歪みに敏感であることや，各フレームの中心と両端で窓関数の値が大きく異なることを考慮して，通常は隣接する時間フレームを半分以上重ねる．結果として，スペクトログラム行列の成分数が元の信号長の 2 倍以上になり，信号圧縮には不向きとなる．他にも，スペクトログラム行列を加工しても，順変換の値域集合から外れると逆変換時に振幅や位相が大きくずれてしまう恐れがある．

変換後の行列の成分数を元の信号の長さとほぼ同じ[1]に保つ方法として，MP3 や AAC などの音信号符号化方式で使用されている，修正離散コサイン変換 (Modified Discrete Cosine Transform: MDCT) [8] がある．離散 STFT のスペクトログラム行列は複素数値だが，MDCT 後の行列は実数値となる．MDCT の逆変換も順変換と同様の計算で実現でき，更に離散 STFT の場合と違い，先頭フレームと最終フレーム以外では逆変換後も行列の所望の成分値を必ず保持できる．一方で，離散 STFT の振幅スペクトログラムと比べて MDCT 後の行列には，信号の平行移動により値が大きく変動するという問題や，複素数値信号の解析には適さないという問題がある．MDCT を拡張した，変調複素重複変換 (Modulated Complex Lapped Transform: MCLT) [9] も提案されているが，変換後の行列は複素数値であり，成分数が MDCT のときの実質 2 倍となる．また，MCLT は離散 STFT とほぼ同等であることが示せる (脚注 7 参照)．

本研究では，スペクトログラムの行列表現が過度に冗長になることを防ぐために，各時間フレームでフレーム長の半分の数の周波数成分のみを求める，周波数間引き STFT (Frequency-Undersampled STFT: FUSTFT) を提案する．実は，Stanković が文献 [10] で FUSTFT と同様の考え方を既に提案しており，これは FUSTFT の特殊な場合 (式 (17) で $\xi = \frac{L_w}{2}$ とした場合) と一致する．FUSTFT を利用すれば，実数値信号と複素数値信号の両方に対して，ほぼ非冗長な

本研究の一部は JSPS 科研費 (19K20361) の御支援を受けて行われた．

[1]先頭フレームと最終フレームでのみ冗長性が生じる (信号長の 2 倍が窓長で割り切れない場合は，最終フレームの直前でも冗長性が生じる)．

スペクトログラムを簡単に与えることができる．一方で，逆変換計算は離散 STFT の場合と異なり簡単ではない [7]．本研究では，文献 [10] に基づいて FUSTFT の 2 種類の逆変換を提示し，これらを三重対角型の線形方程式に帰着させる．そして，離散時間窓関数の形状を対称にしておくことで，帰着した方程式の解が常に一意に定まることを示し，数値実験で FUSTFT と逆変換の正当性を確認する．

2　本稿での短時間フーリエ変換 (STFT) の定義と逆変換

実数全体の集合を $\mathbb{R}$ で，非負実数全体の集合を $\mathbb{R}_+$ で，複素数全体の集合を $\mathbb{C}$ で表し，虚数単位を $\imath \in \mathbb{C}$ で表す，即ち $\imath^2 = -1$ である．ベクトルと行列をそれぞれボールド体の小文字と大文字で表す．ベクトル及び行列の転置と実線形写像の随伴作用素を $(\cdot)^{\mathrm{T}}$ で，行列のエルミート転置と複素線形写像の随伴作用素を $(\cdot)^{\mathrm{H}}$ で，正則行列の逆行列を $(\cdot)^{-1}$ で，写像の合成を $\circ$ で表す．床関数と天井関数をそれぞれ $\lfloor\cdot\rfloor$ と $\lceil\cdot\rceil$ で表す．正実数 a に対して関数 $\mathrm{mod}_a : \mathbb{R} \to [0, a)$ を $\mathrm{mod}_a(b) := b - \left\lfloor \frac{b}{a} \right\rfloor a$ のように定義する．

2.1　連続時間 STFT・離散時間 STFT・離散 STFT

本研究では，非負実数値窓関数 $w : \mathbb{R} \to \mathbb{R}_+$ を用いて[2]，実数値または複素数値信号 $x : \mathbb{R} \to \mathbb{C}$ の連続時間 STFT を

$$X(f,t) = \int_{-\infty}^{\infty} x(\tau) w(\tau - t) e^{-\imath 2\pi f (\tau - t - \frac{T_{\mathrm{s}}}{2})}\, \mathrm{d}\tau \tag{1}$$

のように定義し[3]，離散時間 STFT を

$$X_{\mathrm{d}}(f,t) = \sum_{\tau=-\infty}^{\infty} x(\tau T_{\mathrm{s}}) w(\tau T_{\mathrm{s}} - t) e^{-\imath 2\pi f (\tau T_{\mathrm{s}} - t - \frac{T_{\mathrm{s}}}{2})} \tag{2}$$

$$= \frac{1}{T_{\mathrm{s}}} \sum_{\kappa=-\infty}^{\infty} X(f - \kappa f_{\mathrm{s}}, t) \tag{3}$$

のように定義する[4]．ここで，$f \in \mathbb{R}$，$t \in \mathbb{R}$ であり，$T_{\mathrm{s}} > 0$ は離散時間信号 $x[\tau] := x(\tau T_{\mathrm{s}})$ のサンプリング間隔，$f_{\mathrm{s}} = \frac{1}{T_{\mathrm{s}}}$ はサンプリング周波数である．式 (3) から，$X_{\mathrm{d}}(f,t)$ は f に関して周期 f_{s} の周期関数となるため，離散時間 STFT では $f \in [-\frac{f_{\mathrm{s}}}{2}, \frac{f_{\mathrm{s}}}{2})$ または $f \in [0, f_{\mathrm{s}})$ であるとしてもよい．

以降では L_w を 2 以上の整数とし，窓関数 $w(t)$ は長さ $L_w T_{\mathrm{s}}$ のコンパクトな台を持ち，$t \in (0, L_w T_{\mathrm{s}})$ のときのみ $w(t) > 0$ であり，それ以外では $w(t) = 0$ とする．更に，$w(t)$ の形状は対称かつ単峰で，$w(\frac{L_w T_{\mathrm{s}}}{2} - t) = w(\frac{L_w T_{\mathrm{s}}}{2} + t)$ $(\forall t \in \mathbb{R})$ を満たし，区間 $(0, \frac{L_w T_{\mathrm{s}}}{2}]$ で狭義単調増加[5]とする．

[2]後述の $w(t)$ の条件を考慮して，窓関数を非負実数値に限定している．

[3]多くの文献で複素正弦波を $e^{-\imath 2\pi f(\tau - t)}$ として連続時間 STFT が定義されているが，この定義では後述の離散 STFT を考えた際に，例え窓関数 $w(t)$ の形状が対称であったとしても，離散時間窓関数 $w[\tau]$ の形状は非対称になってしまう．本研究では，$w[\tau]$ も対称になるように，式 (1) でサンプリング間隔 T_{s} も用いて連続時間 STFT を定義している．

[4]多くの文献で式 (2) の時点で離散時間 STFT の引数 t を離散化しているが，$x(t)$ のサンプリング間隔 T_{s} に依らず，窓関数と複素正弦波は任意の時刻 t で計算できるため，式 (2) は $t \in \mathbb{R}$ に対して定義できる．

[5]矩形窓 ($t \in (0, L_w T_{\mathrm{s}})$ で $w(t) = 1$，それ以外で $w(t) = 0$) は，区間 $(0, \frac{L_w T_{\mathrm{s}}}{2}]$ での狭義単調増加性を満たさないため，議論から除外される．

この仮定の下では，式 (1) の連続時間 STFT は

$$X(f,t) = \int_{0}^{L_w T_{\mathrm{s}}} x(\tau + t) w(\tau) e^{-\imath 2\pi f (\tau - \frac{T_{\mathrm{s}}}{2})}\, \mathrm{d}\tau \tag{4}$$

となる．式 (2) の離散時間 STFT は任意の時刻 $t \in \mathbb{R}$ に対して計算可能だが，T_{s} より短い間隔で計算を行う必要はない．整数 l と正整数のシフト量 $\xi\ (\leq L_w)$ を用いて時刻を $t = (l\xi - \frac{1}{2}) T_{\mathrm{s}}$ のように離散化すると，離散時間 STFT は

$$X_{\mathrm{d}}(f, (l\xi - \tfrac{1}{2}) T_{\mathrm{s}}) = \sum_{\tau=0}^{L_w - 1} x[\tau + l\xi] w[\tau] e^{-\imath 2\pi f \tau T_{\mathrm{s}}} \tag{5}$$

となる．ここで，$f \in [0, f_{\mathrm{s}})$，$w[\tau] := w((\tau + \frac{1}{2}) T_{\mathrm{s}})$ である．

次に，式 (5) の離散時間 STFT の周波数 f を離散化する．長さ L_w の離散時間信号から得られる独立な周波数成分の数は L_w 個であることから，非負整数 k を用いて周波数を $f = \frac{k}{L_w} f_{\mathrm{s}}$ と離散化する．長さ $L_x\ (> L_w)$ の離散時間信号 $\boldsymbol{x} := (x[0], x[1], \ldots, x[L_x - 1])^{\mathrm{T}} \in \mathbb{C}^{L_x}$ の離散 STFT を

$$\begin{aligned} \mathrm{STFT}(\boldsymbol{x})[k, l] &= X_{\mathrm{d}}(\tfrac{k}{L_w} f_{\mathrm{s}}, (l\xi - \tfrac{1}{2}) T_{\mathrm{s}}) \\ &= \sum_{\tau=0}^{L_w - 1} x[\tau + l\xi] w[\tau] e^{-\imath 2\pi \frac{k}{L_w} \tau} \end{aligned} \tag{6}$$

のように定義する[6]．ここで，$k = 0, 1, \ldots, L_w - 1$，$l = 0, 1, \ldots, \left\lceil \frac{L_x - L_w}{\xi} \right\rceil$ であり，更に時刻 $t \in (-\infty, -\frac{T_{\mathrm{s}}}{2}]$ と $t \in$

[6]振幅値のみならず位相情報も考慮したい場合には，連続時間 STFT を

$$X(f,t) = \int_{-\infty}^{\infty} x(\tau) w(\tau - t) e^{-\imath 2\pi f \tau}\, \mathrm{d}\tau \tag{7}$$

と定義する．窓関数 $w(t)$ に式 (4) と同様の性質を仮定すると，式 (7) は

$$X(f,t) = \int_{0}^{L_w T_{\mathrm{s}}} x(\tau + t) w(\tau) e^{-\imath 2\pi f (\tau + t)}\, \mathrm{d}\tau \tag{8}$$

となる．離散時間 STFT の引数を離散化することで，$\mathrm{STFT}(\boldsymbol{x})[k, l]$ を

$$X_{\mathrm{d}}(\tfrac{k}{L_w} f_{\mathrm{s}}, (l\xi - \tfrac{1}{2}) T_{\mathrm{s}}) = \sum_{\tau=0}^{L_w - 1} x[\tau + l\xi] w[\tau] e^{-\imath 2\pi \frac{k}{L_w} (\tau + l\xi)} \tag{9}$$

と定義する．式 (6) と式 (9) を比べると，振幅は等しく位相のみが異なり，式 (9) では FFT 後に位相の回転処理が必要となる．文献 [5] で，式 (9) の複素スペクトログラムの方がより低ランクになることが示されている．

ウェーブレット変換のように，L_w の値が異なる複数の窓関数を用いて解析を行う場合には，式 (7) で $w(t)$ の台を $(-\frac{L_w T_{\mathrm{s}}}{2}, \frac{L_w T_{\mathrm{s}}}{2})$ とした

$$X(f,t) = \int_{-\frac{L_w T_{\mathrm{s}}}{2}}^{\frac{L_w T_{\mathrm{s}}}{2}} x(\tau + t) w(\tau) e^{-\imath 2\pi f (\tau + t)}\, \mathrm{d}\tau \tag{10}$$

を用いた方がよい．式 (8) と式 (10) の違いは，時間フレームを開始点で揃えるか，中心で揃えるかであり，L_w が異なる場合は後者の方が良質な解析を行える．式 (10) において L_w が偶数のときは，$\mathrm{STFT}(\boldsymbol{x})[k, l]$ を

$$X_{\mathrm{d}}(\tfrac{k}{L_w} f_{\mathrm{s}}, (l\xi - \tfrac{1}{2}) T_{\mathrm{s}}) = \sum_{\tau=-\frac{L_w}{2}}^{\frac{L_w}{2} - 1} x[\tau + l\xi] w[\tau] e^{-\imath 2\pi \frac{k}{L_w} (\tau + l\xi)} \tag{11}$$

と定義する．一方で，L_w が奇数のときには，$\mathrm{STFT}(\boldsymbol{x})[k, l]$ を

$$X_{\mathrm{d}}(\tfrac{k}{L_w} f_{\mathrm{s}}, l\xi T_{\mathrm{s}}) = \sum_{\tau=-\frac{L_w - 1}{2}}^{\frac{L_w - 1}{2}} x[\tau + l\xi] w[\tau] e^{-\imath 2\pi \frac{k}{L_w} (\tau + l\xi)} \tag{12}$$

と定義する．式 (12) でのみ，$w[\tau] := w(\tau T_{\mathrm{s}})$ である．式 (11) と式 (12) では，$X_{\mathrm{d}}(f, t)$ 及び $w(t)$ の時刻 t が $\frac{T_{\mathrm{s}}}{2}$ ずれているので注意されたい．

$[(L_x-\frac{1}{2})T_\mathrm{s},\infty)$ では $x(t)=0$ であると仮定し，$\boldsymbol{x}$ の末尾に $\lceil\frac{L_x-L_w+\xi}{\xi}\rceil\xi-(L_x-L_w+\xi)$ 個の 0 を埋め込んでいる．式 (6) は，「信号 $\boldsymbol{x}$ の一部を抽出した後に窓関数 $w[\tau]$ を乗じて FFT を行う」という処理で計算でき，L_w が極端に大きくなく，ξ も極端に小さくない限り，複素スペクトログラム $\mathrm{STFT}(\boldsymbol{x})=(\mathrm{STFT}(\boldsymbol{x})[k,l])\in\mathbb{C}^{L_w\times\lceil\frac{L_x-L_w+\xi}{\xi}\rceil}$ を高速に求められる．式 (6) は，窓付き離散フーリエ変換や，離散 Gabor 変換と呼ばれることもあるが，本稿ではこれらの変換と離散 STFT を特に区別せずに同一視する．

各時間フレームで，長さ L_w の離散時間信号から L_w+1 個以上の周波数成分を計算することも可能である．本稿では，これを周波数冗長化 STFT (Frequency-Oversampled STFT: FOSTFT) と呼ぶ．具体的には，N_z を正整数として式 (5) で $f=\frac{k}{L_w+N_\mathrm{z}}f_\mathrm{s}$ と離散化することで，FOSTFT を

$$\begin{aligned}\mathrm{FOSTFT}(\boldsymbol{x})[k,l]&=X_\mathrm{d}(\tfrac{k}{L_w+N_\mathrm{z}}f_\mathrm{s},(l\xi-\tfrac{1}{2})T_\mathrm{s})\\&=\sum_{\tau=0}^{L_w-1}x[\tau+l\xi]w[\tau]e^{-\imath 2\pi\frac{k}{L_w+N_\mathrm{z}}\tau}\end{aligned}\tag{13}$$

と定義する．ここで，$k=0,1,\ldots,L_w+N_\mathrm{z}-1$，$l=0,1,\ldots,\lceil\frac{L_x-L_w}{\xi}\rceil$ である．式 (13) は，各時間フレームの末尾に 0 を N_z 個埋め込んだ後に FFT を行えば計算できる．

2.2 離散 STFT の疑似逆行列に基づく逆変換

離散 STFT の値域集合を $\mathcal{R}:=\{\boldsymbol{X}\in\mathbb{C}^{L_w\times\lceil\frac{L_x-L_w+\xi}{\xi}\rceil}\mid\exists\boldsymbol{x}\in\mathbb{C}^{L_x}\ \boldsymbol{X}=\mathrm{STFT}(\boldsymbol{x})\}$ で表すと，値域に属する複素スペクトログラム $\boldsymbol{X}\in\mathcal{R}$ を対応する離散時間信号 $\boldsymbol{x}$ に戻す線形写像は無数に存在する．値域に含まれない $\boldsymbol{X}\in\mathbb{C}^{L_w\times\lceil\frac{L_x-L_w+\xi}{\xi}\rceil}\setminus\mathcal{R}$ に対しても，最も整合性の取れた $\boldsymbol{x}$ を与えるために，式 (6) の逆変換 (Inverse STFT: ISTFT) を

$$\mathrm{ISTFT}(\boldsymbol{X})=\operatorname*{argmin}_{\boldsymbol{x}\in\mathbb{C}^{L_x}}\|\boldsymbol{X}-\mathrm{STFT}(\boldsymbol{x})\|_\mathrm{F}^2\tag{14}$$

のように定義する．ここで，$\|\cdot\|_\mathrm{F}$ はフロベニウスノルムである．式 (6) の線形写像を $\mathcal{S}:\mathbb{C}^{L_x}\to\mathbb{C}^{L_w\times\lceil\frac{L_x-L_w+\xi}{\xi}\rceil}$ で表すと，式 (14) は「疑似逆行列」の演算に相当するため，$\mathrm{ISTFT}(\boldsymbol{X})=(\mathcal{S}^\mathrm{H}\circ\mathcal{S})^{-1}\circ\mathcal{S}^\mathrm{H}(\boldsymbol{X})$ と表せる．このとき，$\mathcal{S}^\mathrm{H}\circ\mathcal{S}$ は対角行列になるため ISTFT が瞬時に計算できる．式 (13) を $\mathcal{S}$ とした場合も同様の逆変換計算が可能となる．

逆変換の際に現れる対角行列 $\mathcal{S}^\mathrm{H}\circ\mathcal{S}$ の対角成分を観察すると，式 (6) で時刻 t を少しシフトして，離散 STFT を

$$\begin{aligned}\mathrm{STFT}(\boldsymbol{x})[k,l]&=X_\mathrm{d}(\tfrac{k}{L_w}f_\mathrm{s},(l\xi-L_w+\xi-\tfrac{1}{2})T_\mathrm{s})\\&=\sum_{\tau=0}^{L_w-1}x[\tau+l\xi-L_w+\xi]w[\tau]e^{-\imath 2\pi\frac{k}{L_w}\tau}\end{aligned}\tag{15}$$

と定義し直した方がよいことに気付く．同様に FOSTFT も

$$\begin{aligned}\mathrm{FOSTFT}(\boldsymbol{x})[k,l]&=X_\mathrm{d}(\tfrac{k}{L_w+N_\mathrm{z}}f_\mathrm{s},(l\xi-L_w+\xi-\tfrac{1}{2})T_\mathrm{s})\\&=\sum_{\tau=0}^{L_w-1}x[\tau+l\xi-L_w+\xi]w[\tau]e^{-\imath 2\pi\frac{k}{L_w+N_\mathrm{z}}\tau}\end{aligned}\tag{16}$$

と定義し直す．式 (15) と式 (16) では，k の値は元のままで $l=0,1,\ldots,\lceil\frac{L_x+L_w-2\xi}{\xi}\rceil$ となる．これらの式は $\boldsymbol{x}$ の先端に $L_w-\xi$ 個の 0 を，末尾に $\lceil\frac{L_x+L_w-\xi}{\xi}\rceil\xi-L_x$ 個の 0 を埋め込んだ後に式 (6) と式 (13) を適用することに等しい．式 (15) の写像を $\mathcal{S}:\mathbb{C}^{L_x}\to\mathbb{C}^{L_w\times\lceil\frac{L_x+L_w-\xi}{\xi}\rceil}$ で表すと，対角行列 $\mathcal{S}^\mathrm{H}\circ\mathcal{S}$ の対角成分が周期 ξ の周期信号となり，逆変換が更に容易になる [7]．式 (16) の場合も同様である．

3 周波数間引き STFT

人間の聴覚が音信号のブロック歪みに敏感であることや，時間フレームの中心と両端で窓関数の値が大きく異なることなどから，式 (15) の離散 STFT や式 (16) の FOSTFT でシフト量を $\xi=L_w$ のように設定することはまずなく，$\xi\le\frac{L_w}{2}$ のように設定する場合がほとんどである．しかし，この場合にはスペクトログラム行列の成分数が信号長 L_x の 2 倍以上となり，信号圧縮には向かない．また，スペクトログラムを所望の値に加工しても，値域集合から外れると逆変換時に振幅や位相がずれてしまう問題も存在した．

本研究では，より効率的な時間周波数解析の実現のため，4 の倍数の長さ L_w の離散時間信号から $\frac{L_w}{2}$ 個の周波数成分を計算する，周波数間引き STFT (Frequency-Undersampled STFT: FUSTFT) を提案する．式 (5) で $\xi\le\frac{L_w}{2}$ とし周波数を $f=\frac{2k}{L_w}f_\mathrm{s}$ と離散化することで，タイプ I の FUSTFT を

$$\begin{aligned}\mathrm{FUSTFT}_\mathrm{I}(\boldsymbol{x})[k,l]&=X_\mathrm{d}(\tfrac{2k}{L_w}f_\mathrm{s},(l\xi-L_w+\xi-\tfrac{1}{2})T_\mathrm{s})\\&=\sum_{\tau=0}^{L_w-1}x[\tau+l\xi-L_w+\xi]w[\tau]e^{-\imath 2\pi\frac{2k}{L_w}\tau}\end{aligned}\tag{17}$$

と定義する．$f=\frac{2k+1}{L_w}f_\mathrm{s}$ として，タイプ II の FUSTFT を

$$\begin{aligned}\mathrm{FUSTFT}_\mathrm{II}(\boldsymbol{x})[k,l]&=X_\mathrm{d}(\tfrac{2k+1}{L_w}f_\mathrm{s},(l\xi-L_w+\xi-\tfrac{1}{2})T_\mathrm{s})\\&=\sum_{\tau=0}^{L_w-1}x[\tau+l\xi-L_w+\xi]w[\tau]e^{-\imath 2\pi\frac{2k+1}{L_w}\tau}\end{aligned}\tag{18}$$

と定義する．上記を交互に用いて，タイプ III の FUSTFT を

$$\mathrm{FUSTFT}_\mathrm{III}(\boldsymbol{x})[k,l]$$
$$=\begin{cases}\displaystyle\sum_{\tau=0}^{L_w-1}x[\tau+l\xi-L_w+\xi]w[\tau]e^{-\imath 2\pi\frac{2k}{L_w}\tau} & (l\text{ が偶数})\\ \displaystyle\sum_{\tau=0}^{L_w-1}x[\tau+l\xi-L_w+\xi]w[\tau]e^{-\imath 2\pi\frac{2k+1}{L_w}\tau} & (l\text{ が奇数})\end{cases}\tag{19}$$

と定義する．式 (17) から式 (19) では，$k=0,1,\ldots,\frac{L_w}{2}-1$，$l=0,1,\ldots,\lceil\frac{L_x+L_w-2\xi}{\xi}\rceil$ であり，複素スペクトログラムの周波数成分の数が式 (15) の場合[7]と比べて半分になる．

[7] 式 (16) で $N_\mathrm{z}=L_w$ とした後に式 (18) を利用して，離散 STFT を

$$\mathrm{STFT}(\boldsymbol{x})[k,l]=\sum_{\tau=0}^{L_w-1}x[\tau+l\xi-L_w+\xi]w[\tau]e^{-\imath 2\pi\frac{2k+1}{2L_w}\tau}\tag{20}$$

と定義し直すことも可能である．式 (20) の線形写像を $\mathcal{S}$ とした場合も，逆変換 $(\mathcal{S}^\mathrm{H}\circ\mathcal{S})^{-1}\circ\mathcal{S}^\mathrm{H}$ が式 (15) の場合と同様に計算できる．MCLT [9] と式 (20) で $\xi=\frac{L_w}{2}$ とした結果では，振幅は等しく位相のみが異なる．

4 周波数間引き STFT の 2 種類の逆変換

4.1 疑似逆行列に基づく逆変換

式 (17) から式 (19) のいずれかの線形写像を $\mathcal{S}:\mathbb{C}^{L_x}\to\mathbb{C}^{\frac{L_w}{2}\times\lceil\frac{L_x+L_w-\xi}{\xi}\rceil}$ で表すと，疑似逆行列に基づく逆変換は $(\mathcal{S}^{\mathrm{H}}\circ\mathcal{S})^{-1}\circ\mathcal{S}^{\mathrm{H}}$ となる．FUSTFT の場合には，$\mathcal{S}^{\mathrm{H}}\circ\mathcal{S}$ が

$$\begin{bmatrix}
a_0 & & & b_0 & & & & & & & \\
& a_1 & & & b_1 & & & & & & \\
& & \ddots & & & \ddots & & & & & \\
b_0 & & & a_{\frac{L_w}{2}} & & & b_{\frac{L_w}{2}} & & & & \\
& b_1 & & & a_{\frac{L_w}{2}+1} & & & \ddots & & & \\
& & \ddots & & & \ddots & & & \ddots & & \\
& & & b_{\frac{L_w}{2}} & & & \ddots & & & \ddots & \\
& & & & \ddots & & & a_{L_x-\frac{L_w}{2}-1} & & & b_{L_x-\frac{L_w}{2}-1} \\
& & & & & \ddots & & & \ddots & & \\
& & & & & & \ddots & & & \ddots & \\
& & & & & & & b_{L_x-\frac{L_w}{2}-1} & & & a_{L_x-1}
\end{bmatrix}$$

のような行列[8]となる．対角成分は全てのタイプにおいて，

$$a_i=\frac{L_w}{2}\sum_{l=0}^{\lceil\frac{L_w}{\xi}\rceil-1}w^2[m_i+l\xi] \tag{21}$$

となる．ここで，$m_i=\mathrm{mod}_\xi(i+L_w)$ であり，$\tau\geq L_w$ のとき $w[\tau]=0$ とした．非零の非対角成分は，タイプ I で

$$b_i=\frac{L_w}{2}\sum_{l=0}^{\lceil\frac{L_w}{2\xi}\rceil-1}w[m_i+l\xi]w[m_i+l\xi+\tfrac{L_w}{2}] \tag{22}$$

となり，タイプ II では符号が反転して

$$b_i=-\frac{L_w}{2}\sum_{l=0}^{\lceil\frac{L_w}{2\xi}\rceil-1}w[m_i+l\xi]w[m_i+l\xi+\tfrac{L_w}{2}] \tag{23}$$

となる．タイプ III の場合は少し複雑だが

$$b_i=\frac{L_w}{2}\sum_{l=0}^{\lceil\frac{L_w}{2\xi}\rceil-1}(-1)^{\lfloor\frac{i+L_w-\xi}{\xi}\rfloor+l}w[m_i+l\xi]w[m_i+l\xi+\tfrac{L_w}{2}] \tag{24}$$

となる．式 (21) の a_i と式 (22) 及び式 (23) の b_i は周期 ξ の周期信号に，式 (24) の b_i は周期 2ξ の周期信号になる．

与えられた複素スペクトログラム $\boldsymbol{X}\in\mathbb{C}^{\frac{L_w}{2}\times\lceil\frac{L_x+L_w-\xi}{\xi}\rceil}$ に対して，$\mathcal{S}^{\mathrm{H}}(\boldsymbol{X})=:\boldsymbol{y}=:(y[0],y[1],\ldots,y[L_x-1])^{\mathrm{T}}$ と定義すれば，線形方程式 $(\mathcal{S}^{\mathrm{H}}\circ\mathcal{S})\boldsymbol{x}=\boldsymbol{y}$ の解 $\boldsymbol{x}$ が，$\boldsymbol{X}$ の逆変換結果となる．この方程式は，$\frac{L_w}{2}$ 個の独立な方程式

$$\begin{bmatrix}
a_0^{\langle i\rangle} & b_0^{\langle i\rangle} & & & \\
b_0^{\langle i\rangle} & a_1^{\langle i\rangle} & b_1^{\langle i\rangle} & & \\
& \ddots & \ddots & \ddots & \\
& & b_{n_i-3}^{\langle i\rangle} & a_{n_i-2}^{\langle i\rangle} & b_{n_i-2}^{\langle i\rangle} \\
& & & b_{n_i-2}^{\langle i\rangle} & a_{n_i-1}^{\langle i\rangle}
\end{bmatrix}
\begin{bmatrix}
x[i] \\ x[i+\frac{L_w}{2}] \\ \vdots \\ x[i+\frac{(n_i-2)L_w}{2}] \\ x[i+\frac{(n_i-1)L_w}{2}]
\end{bmatrix}
=
\begin{bmatrix}
y[i] \\ y[i+\frac{L_w}{2}] \\ \vdots \\ y[i+\frac{(n_i-2)L_w}{2}] \\ y[i+\frac{(n_i-1)L_w}{2}]
\end{bmatrix} \tag{25}$$

に帰着される．ここで，$i=0,1,\ldots,\frac{L_w}{2}-1$ であり，$n_i=\lceil\frac{2(L_x-i)}{L_w}\rceil$，$a_j^{\langle i\rangle}:=a_{i+j\frac{L_w}{2}}$，$b_j^{\langle i\rangle}:=b_{i+j\frac{L_w}{2}}$ である．式 (25) の左辺の行列は三重対角行列であり，LU 分解[9]が $\mathcal{O}(n)$ で求まる [11]．このため，逆変換結果 $\boldsymbol{x}$ が高速に得られる．

特に，タイプ I かタイプ II で $\mathrm{mod}_\xi(\frac{L_w}{2})=0$ のときや，もしくはタイプ III で $\mathrm{mod}_{2\xi}(\frac{L_w}{2})=0$ のときは，$a_0^{\langle i\rangle}=a_1^{\langle i\rangle}=\cdots=a_{n_i-1}^{\langle i\rangle}=a_i$，$b_0^{\langle i\rangle}=b_1^{\langle i\rangle}=\cdots=b_{n_i-2}^{\langle i\rangle}=b_i$ であるため，式 (25) の左辺の行列は三重対角テプリッツ行列となり，固有値分解も閉形式で与えられる．固有値は $\lambda_q^{\langle i\rangle}=a_i+2b_i\cos(\frac{q\pi}{n_i+1})>a_i-2|b_i|>0$，固有ベクトルは $\boldsymbol{u}_q=\sqrt{\frac{2}{n_i+1}}(\sin(\frac{q\pi}{n_i+1}),\sin(\frac{2q\pi}{n_i+1}),\ldots,\sin(\frac{n_iq\pi}{n_i+1}))^{\mathrm{T}}\in\mathbb{R}^{n_i}$ $(q=1,2,\ldots,n_i)$ となる [12]．即ち，この場合はタイプ I の離散サイン変換 [13] を用いても式 (25) の解が得られる．

タイプ III で $\mathrm{mod}_\xi(\frac{L_w}{2})=0$ かつ $\mathrm{mod}_{2\xi}(\frac{L_w}{2})\neq0$ のときには，$b_j^{\langle i\rangle}=(-1)^jb_i$ となる．この場合も式 (25) の左辺の行列の固有値分解を閉形式で与えられる．固有値は $\lambda_q^{\langle i\rangle}=a_i+2b_i\cos(\frac{q\pi}{n_i+1})>0$ で，固有ベクトルは $\boldsymbol{u}_q=\sqrt{\frac{2}{n_i+1}}\times(\sin(\frac{q\pi}{n_i+1}),\sin(\frac{2q\pi}{n_i+1}),-\sin(\frac{3q\pi}{n_i+1}),-\sin(\frac{4q\pi}{n_i+1}),\sin(\frac{5q\pi}{n_i+1}),\ldots,(-1)^{\frac{(n_i-2)(n_i-1)}{2}}\sin(\frac{n_iq\pi}{n_i+1}))^{\mathrm{T}}\in\mathbb{R}^{n_i}$ である．この場合は，タイプ I の離散サイン変換と適当な符号反転処理を施すことでも式 (25) の解を得ることが可能となる．

4.2 周期条件を仮定した際の疑似逆行列に基づく逆変換

まず，p を $\mathrm{mod}_{\frac{L_w}{2}}(\lceil\frac{L_x+L_w-\xi}{\xi}\rceil\xi+p\xi)=0$ を満たす最小の非負整数[10]とし，$L_{\mathrm{p}}:=\lceil\frac{L_x+L_w-\xi}{\xi}\rceil\xi+p\xi>L_x$ とする．$\boldsymbol{x}_{\mathrm{p}}:=(x[0],x[1],\ldots,x[L_x-1],x[L_x],\ldots,x[L_{\mathrm{p}}-1])^{\mathrm{T}}:=(\boldsymbol{x}^{\mathrm{T}},\boldsymbol{0}_{L_{\mathrm{p}}-L_x}^{\mathrm{T}})^{\mathrm{T}}\in\mathbb{C}^{L_{\mathrm{p}}}$ と定義し，式 (17) から式 (19) を $l=0,1,\ldots,\lceil\frac{L_x+L_w-2\xi}{\xi}\rceil+p$ に対して計算する線形写像を $\mathcal{S}_{\mathrm{p}}:\mathbb{C}^{L_{\mathrm{p}}}\to\mathbb{C}^{\frac{L_w}{2}\times(\lceil\frac{L_x+L_w-\xi}{\xi}\rceil+p)}$ とする．ただし，$\mathcal{S}_{\mathrm{p}}$ は $\tau<0$ のとき周期条件 $x[\tau]=x[L_{\mathrm{p}}+\tau]$ を仮定して計算しているものとする．複素スペクトログラム $\boldsymbol{X}\in\mathbb{C}^{\frac{L_w}{2}\times\lceil\frac{L_x+L_w-\xi}{\xi}\rceil}$ から $\boldsymbol{X}_{\mathrm{p}}:=[\boldsymbol{X},\boldsymbol{O}_{\frac{L_w}{2}\times p}]\in\mathbb{C}^{\frac{L_w}{2}\times(\lceil\frac{L_x+L_w-\xi}{\xi}\rceil+p)}$ を定義し，逆変換 $\boldsymbol{x}_{\mathrm{p}}=(\mathcal{S}_{\mathrm{p}}^{\mathrm{H}}\circ\mathcal{S}_{\mathrm{p}})^{-1}\circ\mathcal{S}_{\mathrm{p}}^{\mathrm{H}}(\boldsymbol{X}_{\mathrm{p}})$ を計算する．$\mathcal{S}_{\mathrm{p}}^{\mathrm{H}}\circ\mathcal{S}_{\mathrm{p}}$ は

$$\begin{bmatrix}
a_0 & & & b_0 & & & & & & b_{L_{\mathrm{p}}-\frac{L_w}{2}} & & \\
& a_1 & & & b_1 & & & & & & \ddots & \\
& & \ddots & & & \ddots & & & & & & b_{L_{\mathrm{p}}-1} \\
b_0 & & & a_{\frac{L_w}{2}} & & & b_{\frac{L_w}{2}} & & & & & \\
& b_1 & & & a_{\frac{L_w}{2}+1} & & & \ddots & & & & \\
& & \ddots & & & \ddots & & & \ddots & & & \\
& & & b_{\frac{L_w}{2}} & & & \ddots & & & \ddots & & \\
& & & & \ddots & & & a_{L_{\mathrm{p}}-\frac{L_w}{2}-1} & & & b_{L_{\mathrm{p}}-\frac{L_w}{2}-1} & \\
b_{L_{\mathrm{p}}-\frac{L_w}{2}} & & & & & \ddots & & & \ddots & & & \\
& \ddots & & & & & \ddots & & & \ddots & & \\
& & b_{L_{\mathrm{p}}-1} & & & & & b_{L_{\mathrm{p}}-\frac{L_w}{2}-1} & & & a_{L_{\mathrm{p}}-1} &
\end{bmatrix}$$

---

[8] 第 l フレーム抽出行列 $\boldsymbol{P}_l\in\mathbb{R}_+^{L_w\times L_x}$，窓行列 $\boldsymbol{W}=\mathrm{diag}(w[\tau])\in\mathbb{R}_+^{L_w\times L_w}$，タイプ I の間引き離散フーリエ変換行列 $\boldsymbol{F}_{\mathrm{u}}\in\mathbb{C}^{\frac{L_w}{2}\times L_w}$ を用いれば，タイプ I の場合は $\mathcal{S}^{\mathrm{H}}\circ\mathcal{S}=\sum_l\boldsymbol{P}_l^{\mathrm{T}}\boldsymbol{W}\boldsymbol{F}_{\mathrm{u}}^{\mathrm{H}}\boldsymbol{F}_{\mathrm{u}}\boldsymbol{W}\boldsymbol{P}_l$ である．

[9] 三重対角行列の LU 分解は，Thomas Algorithm とも呼ばれている．

[10] この条件はタイプ I かタイプ II の FUSTFT の場合である．タイプ III の場合は，p は更に $\mathrm{mod}_2(\lceil\frac{L_x+L_w-\xi}{\xi}\rceil+p)=0$ も満たす必要がある．

のような行列となる．ここで，a_i と b_i は式 (21) から式 (24) のものと同じである．$\boldsymbol{x}_\mathrm{p}$ を計算した後に先頭の L_x 個の成分を抽出することで，最終的な逆変換結果 $\boldsymbol{x}$ を得る．

$\mathcal{S}_\mathrm{p}^\mathrm{H}(\boldsymbol{X}_\mathrm{p}) =: \boldsymbol{y}_\mathrm{p} =: (y[0], y[1], \ldots, y[L_\mathrm{p}-1])^\mathrm{T}$ を定義し，線形方程式 $(\mathcal{S}_\mathrm{p}^\mathrm{H} \circ \mathcal{S}_\mathrm{p})\boldsymbol{x}_\mathrm{p} = \boldsymbol{y}_\mathrm{p}$ の解 $\boldsymbol{x}_\mathrm{p}$ を計算すればよい．この線形方程式は，$\frac{L_w}{2}$ 個のサイズの等しい独立な方程式

$$\begin{bmatrix} a_0^{\langle i\rangle} & b_0^{\langle i\rangle} & & & b_{n_i-1}^{\langle i\rangle} \\ b_0^{\langle i\rangle} & a_1^{\langle i\rangle} & b_1^{\langle i\rangle} & & \\ & \ddots & \ddots & \ddots & \\ & & b_{n_i-3}^{\langle i\rangle} & a_{n_i-2}^{\langle i\rangle} & b_{n_i-2}^{\langle i\rangle} \\ b_{n_i-1}^{\langle i\rangle} & & & b_{n_i-2}^{\langle i\rangle} & a_{n_i-1}^{\langle i\rangle} \end{bmatrix} \begin{bmatrix} x[i] \\ x[i+\frac{L_w}{2}] \\ \vdots \\ x[i+\frac{(n_i-2)L_w}{2}] \\ x[i+\frac{(n_i-1)L_w}{2}] \end{bmatrix} = \begin{bmatrix} y[i] \\ y[i+\frac{L_w}{2}] \\ \vdots \\ y[i+\frac{(n_i-2)L_w}{2}] \\ y[i+\frac{(n_i-1)L_w}{2}] \end{bmatrix} \tag{26}$$

に帰着される[11]．ここで，$i = 0, 1, \ldots, \frac{L_w}{2}-1$ であり，$n_0 = n_1 = \cdots = n_{\frac{L_w}{2}-1} = \frac{2L_\mathrm{p}}{L_w}$，$a_j^{\langle i\rangle} := a_{i+j\frac{L_w}{2}}$，$b_j^{\langle i\rangle} := b_{i+j\frac{L_w}{2}}$ である．式 (26) の左辺の行列は周期三重対角行列であり，LU 分解が $\mathcal{O}(n)$ で計算できる [14]．結果として，方程式の解 $\boldsymbol{x}_\mathrm{p}$ 及び最終的な逆変換結果 $\boldsymbol{x}$ を高速に求められる．

特に，タイプ I かタイプ II で $\mathrm{mod}_\xi(\frac{L_w}{2}) = 0$ のときや，もしくはタイプ III で $\mathrm{mod}_{2\xi}(\frac{L_w}{2}) = 0$ のときは，$a_0^{\langle i\rangle} = a_1^{\langle i\rangle} = \cdots = a_{n_i-1}^{\langle i\rangle} = a_i$，$b_0^{\langle i\rangle} = b_1^{\langle i\rangle} = \cdots = b_{n_i-1}^{\langle i\rangle} = b_i$ であるため，式 (26) の左辺の行列は巡回行列となり，FFT を用いて対角化 (固有値分解) できる．したがって，固有値は $\lambda_q^{\langle i\rangle} = a_i + 2b_i\cos(\frac{2q\pi}{n_i}) \geq a_i - 2|b_i| > 0$ で[12]，固有ベクトルは $\boldsymbol{u}_q = \frac{1}{\sqrt{n_i}}(1, e^{-\imath\frac{2q\pi}{n_i}}, e^{-\imath\frac{4q\pi}{n_i}}, \ldots, e^{-\imath\frac{2(n_i-1)q\pi}{n_i}})^\mathrm{T} \in \mathbb{C}^{n_i}$ $(q = 0, 1, \ldots, n_i-1)$ である．文献 [10] では，離散時間窓関数を $w[\tau] = w(\tau T_\mathrm{s})$ と定義しているため，$|b_i| = \frac{a_i}{2}$ となり，固有値が $\lambda_q^{\langle i\rangle} = 0$ となってしまう場合があった．本研究では，$w[\tau] = w((\tau+\frac{1}{2})T_\mathrm{s})$ と定義することで，離散時間窓関数の形状が対称となり，$|b_i| < \frac{a_i}{2}$ が常に成立するため，固有値 $\lambda_q^{\langle i\rangle}$ は必ず正の値となる．結果として，FFT を用いても式 (26) の解が常に求まる．ただし，L_w が大きいと悪条件[13]になる可能性が残るので，注意が必要である．

タイプ III で $\mathrm{mod}_\xi(\frac{L_w}{2}) = 0$ かつ $\mathrm{mod}_{2\xi}(\frac{L_w}{2}) \neq 0$ のときには，$b_j^{\langle i\rangle} = (-1)^j b_i$ となる．この場合も式 (26) の左辺の行列の固有値分解を閉形式で与えられるが，$\mathrm{mod}_4(n_i)$ の値に応じて場合分けが必要となる．脚注 10 と $\mathrm{mod}_\xi(\frac{L_w}{2}) = 0$ かつ $\mathrm{mod}_{2\xi}(\frac{L_w}{2}) \neq 0$ から，タイプ III では n_i が偶数となることに注意すれば，$\mathrm{mod}_4(n_i) = 0$ もしくは $\mathrm{mod}_4(n_i) = 2$ の場合を考えれば十分である．$\mathrm{mod}_4(n_i) = 0$ の場合には，第 4.1 節と同じ符号反転処理を行うことにより，固有値は $\lambda_q^{\langle i\rangle} = a_i + 2b_i\cos(\frac{2q\pi}{n_i}) > 0$，固有ベクトルは $\boldsymbol{u}_q = \frac{1}{\sqrt{n_i}} \times (1, e^{-\imath\frac{2q\pi}{n_i}}, -e^{-\imath\frac{4q\pi}{n_i}}, -e^{-\imath\frac{6q\pi}{n_i}}, e^{-\imath\frac{8q\pi}{n_i}}, \ldots, -e^{-\imath\frac{2(n_i-1)q\pi}{n_i}})^\mathrm{T} \in \mathbb{C}^{n_i}$ となる．$\mathrm{mod}_4(n_i) = 2$ の場合は少し複雑だが，固有値は $\lambda_q^{\langle i\rangle} = a_i + 2b_i\sin(\frac{2q\pi}{n_i}) > 0$，固有ベクトルは $\boldsymbol{u}_q = \frac{1}{\sqrt{n_i}}(1, \imath e^{-\imath\frac{2q\pi}{n_i}}, e^{-\imath\frac{4q\pi}{n_i}}, \imath e^{-\imath\frac{6q\pi}{n_i}}, \ldots, \imath e^{-\imath\frac{2(n_i-1)q\pi}{n_i}})^\mathrm{T} \in \mathbb{C}^{n_i}$ となる[14]．以上より，FFT と適当な符号反転処理もしくは $\imath$ の乗算を施すことでも式 (26) の解を得ることができる．

[11] 複素スペクトログラム $\boldsymbol{X}$ に適切なサイズの零行列を結合して $\boldsymbol{X}_\mathrm{p}$ へと拡大することにより，線形方程式が式 (26) のように分解可能となる．

[12] 固有値は $(a_i, b_i, 0, 0, \ldots, 0, b_i)^\mathrm{T} \in \mathbb{R}^{n_i}$ を FFT することで求まる．

[13] 条件数は $\mathrm{cond}(\mathcal{S}_\mathrm{p}^\mathrm{H} \circ \mathcal{S}_\mathrm{p}) = \max_{i,q}\{\lambda_q^{\langle i\rangle}\}/\min_{i,q}\{\lambda_q^{\langle i\rangle}\}$ と表せる．

4.3　2 種類の逆変換の違い

多くの文献で，第 4.2 節のように周期条件を仮定して，離散 STFT を説明している．しかしながら，式 (15) の離散 STFT や式 (16) の FOSTFT の場合は，$\mathcal{S}^\mathrm{H} \circ \mathcal{S}$ と $\mathcal{S}_\mathrm{p}^\mathrm{H} \circ \mathcal{S}_\mathrm{p}$ がどちらも対角行列となる [7]．これにより 2 種類の逆変換結果は常に一致するため，離散 STFT や FOSTFT を説明する際には，周期条件を仮定しなくても特に問題ない[15]．

一方で，FUSTFT の場合には，これら 2 種類の逆変換は異なる性質を持つ．まず FUSTFT を $\mathcal{S}$ で表し，値域集合を $\mathcal{R}_\mathcal{S} := \{\boldsymbol{X} \in \mathbb{C}^{\frac{L_w}{2} \times \lceil \frac{L_x+L_w-\xi}{\xi} \rceil} \mid \exists \boldsymbol{x} \in \mathbb{C}^{L_x}\ \boldsymbol{X} = \mathcal{S}(\boldsymbol{x})\}$ とする．値域に属する複素スペクトログラム $\boldsymbol{X} \in \mathcal{R}_\mathcal{S}$ に対しては，どちらの逆変換も対応する離散時間信号 $\boldsymbol{x}$ を復元できる[16]．値域外にある $\boldsymbol{X} \in \mathbb{C}^{\frac{L_w}{2} \times \lceil \frac{L_x+L_w-\xi}{\xi} \rceil} \setminus \mathcal{R}_\mathcal{S}$ に対しては，$(\mathcal{S}^\mathrm{H} \circ \mathcal{S})^{-1} \circ \mathcal{S}^\mathrm{H}$ を用いれば $\boldsymbol{X}$ と最も整合性の取れた $\boldsymbol{x}$ が得られる．$(\mathcal{S}_\mathrm{p}^\mathrm{H} \circ \mathcal{S}_\mathrm{p})^{-1} \circ \mathcal{S}_\mathrm{p}^\mathrm{H}$ を用いた際には $\boldsymbol{X}_\mathrm{p}$ と最も整合性の取れた $\boldsymbol{x}_\mathrm{p}$ が得られるが，これは $\boldsymbol{x}_\mathrm{p}$ の後半 $L_\mathrm{p} - L_x$ 個の成分が非零となることを意味する．前者の逆変換 $(\mathcal{S}^\mathrm{H} \circ \mathcal{S})^{-1} \circ \mathcal{S}^\mathrm{H}$ を常に使えばよいと思われがちだが，実は $\xi = \frac{L_w}{2}$ のときは後者の逆変換 $(\mathcal{S}_\mathrm{p}^\mathrm{H} \circ \mathcal{S}_\mathrm{p})^{-1} \circ \mathcal{S}_\mathrm{p}^\mathrm{H}$ が特別な性質を持つようになる．$\xi = \frac{L_w}{2}$ のときのみ $\mathcal{S}_\mathrm{p}$ が非冗長な変換となり，任意の $\boldsymbol{X}_\mathrm{p}$ に対して，$\boldsymbol{X}_\mathrm{p} = \mathcal{S}_\mathrm{p}(\boldsymbol{x}_\mathrm{p})$ を満たすような $\boldsymbol{x}_\mathrm{p}$ が逆変換により必ず得られる．結果として，任意の複素スペクトログラム $\boldsymbol{X} \in \mathbb{C}^{\frac{L_w}{2} \times \lceil \frac{L_x+L_w-\xi}{\xi} \rceil}$ に対して，先頭フレームと最終フレーム以外での整合性を完全に保証する離散時間信号 $\boldsymbol{x}$ が常に復元可能となる[17]．

5　数値実験

まず，第 4 節の逆変換により離散時間信号が復元できることを確認する．音信号 $\boldsymbol{x}$ には，約 15 秒の「数字を数える

[14] 以上の議論をまとめると，$\xi = \frac{L_w}{2}$ で周期条件を仮定して逆変換を考えた際，離散時間窓関数を $w[\tau] = w(\tau T_\mathrm{s})$ と定義すると，タイプ I の FUSTFT では n_i が偶数のときに，タイプ II は常に，タイプ III では $\mathrm{mod}_4(n_i) = 0$ のときに，ランク落ちが発生して逆変換が不可能となる．

[15] 実際に，本研究では周期条件を仮定せずに連続時間 STFT から離散時間 STFT を定義し，更に離散 STFT や FOSTFT を矛盾なく導出した．式 (20) の離散 STFT で $\xi = \frac{L_w}{2}$ とした結果の実部と似た構造を持つ，修正離散コサイン変換 (Modified Discrete Cosine Transform: MDCT) [8]

$$\mathrm{MDCT}(\boldsymbol{x})[k,l] = \sum_{\tau=0}^{L_w-1} x[\tau+(l-1)\tfrac{L_w}{2}]\,w[\tau]\cos\left(\tfrac{2k+1}{L_w}(\tau+\tfrac{L_w}{4}+\tfrac{1}{2})\pi\right) \tag{27}$$

も考えてみる．ここで，$k = 0, 1, \ldots, \frac{L_w}{2} - 1$，$l = 0, 1, \ldots, \lceil \frac{2L_x}{L_w} \rceil$ である．式 (27) の MDCT を実線形写像 $\mathcal{C}$ で表すと，$\mathcal{C}^\mathrm{T} \circ \mathcal{C}$ と周期条件を仮定した $\mathcal{C}_\mathrm{p}^\mathrm{T} \circ \mathcal{C}_\mathrm{p}$ はどちらも対角行列となり，2 種類の逆変換結果は一致するが，MDCT では周期条件に触れずに $(\mathcal{C}^\mathrm{T} \circ \mathcal{C})^{-1} \circ \mathcal{C}^\mathrm{T}$ を用いる．$\mathcal{C}^\mathrm{T} \circ \mathcal{C}$ が対角行列になるためには，窓関数 $w[\tau]$ の対称性が必要となる．

[16] $(\mathcal{S}^\mathrm{H} \circ \mathcal{S})^{-1} \circ \mathcal{S}^\mathrm{H}$ の場合は自明である．$(\mathcal{S}_\mathrm{p}^\mathrm{H} \circ \mathcal{S}_\mathrm{p})^{-1} \circ \mathcal{S}_\mathrm{p}^\mathrm{H}$ の場合でも，$\boldsymbol{x}_\mathrm{p}$ の後半 $L_\mathrm{p} - L_x$ 個の成分が 0 になり，前半部分が $\boldsymbol{x}$ になる．

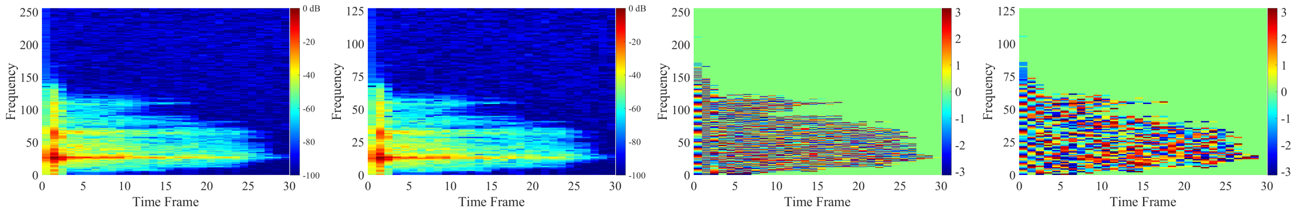
[17] これは MDCT の逆変換 $(\mathcal{C}^\mathrm{T} \circ \mathcal{C})^{-1} \circ \mathcal{C}^\mathrm{T}$ が有する性質と一致する．

表 1: 値域内の複素スペクトログラムからの信号復元結果

L_w	ξ	逆変換	$\frac{\|\boldsymbol{x}-\hat{\boldsymbol{x}}\|_2}{\|\boldsymbol{x}\|_2}$	$\frac{\|\boldsymbol{X}-\mathcal{S}(\hat{\boldsymbol{x}})\|_{\mathrm{F}}}{\|\boldsymbol{X}\|_{\mathrm{F}}}$	$\frac{\|\boldsymbol{X}-\mathcal{S}(\hat{\boldsymbol{x}})\|_{\mathrm{F}}^{\mathrm{int}}}{\|\boldsymbol{X}\|_{\mathrm{F}}^{\mathrm{int}}}$
2,048	$\frac{L_w}{2}$	$\mathcal{S}$	5.0×10^{-14}	8.3×10^{-16}	8.3×10^{-16}
		$\mathcal{S}_{\mathrm{p}}$	1.0×10^{-13}	3.8×10^{-15}	9.6×10^{-16}
	$\frac{L_w}{8}$	$\mathcal{S}$	6.5×10^{-16}	6.7×10^{-16}	6.7×10^{-16}
		$\mathcal{S}_{\mathrm{p}}$	4.7×10^{-16}	5.1×10^{-16}	5.1×10^{-16}
8,192	$\frac{L_w}{2}$	$\mathcal{S}$	1.3×10^{-14}	5.7×10^{-16}	5.7×10^{-16}
		$\mathcal{S}_{\mathrm{p}}$	1.8×10^{-12}	1.1×10^{-13}	9.8×10^{-16}
	$\frac{L_w}{8}$	$\mathcal{S}$	4.4×10^{-16}	5.1×10^{-16}	5.1×10^{-16}
		$\mathcal{S}_{\mathrm{p}}$	4.4×10^{-16}	5.0×10^{-16}	5.0×10^{-16}

表 2: 値域外の複素スペクトログラムからの信号復元結果

L_w	ξ	逆変換	$\frac{\|\boldsymbol{x}-\hat{\boldsymbol{x}}\|_2}{\|\boldsymbol{x}\|_2}$	$\frac{\|\widehat{\boldsymbol{X}}-\mathcal{S}(\hat{\boldsymbol{x}})\|_{\mathrm{F}}}{\|\widehat{\boldsymbol{X}}\|_{\mathrm{F}}}$	$\frac{\|\widehat{\boldsymbol{X}}-\mathcal{S}(\hat{\boldsymbol{x}})\|_{\mathrm{F}}^{\mathrm{int}}}{\|\widehat{\boldsymbol{X}}\|_{\mathrm{F}}^{\mathrm{int}}}$
2,048	$\frac{L_w}{2}$	$\mathcal{S}$	2.9×10^{-2}	6.0×10^{-4}	2.5×10^{-4}
		$\mathcal{S}_{\mathrm{p}}$	3.5×10^{-2}	1.2×10^{-3}	1.7×10^{-15}
	$\frac{L_w}{8}$	$\mathcal{S}$	8.0×10^{-3}	1.3×10^{-2}	1.3×10^{-2}
		$\mathcal{S}_{\mathrm{p}}$	8.0×10^{-3}	1.3×10^{-2}	1.3×10^{-2}
8,192	$\frac{L_w}{2}$	$\mathcal{S}$	2.7×10^{-2}	1.6×10^{-3}	4.2×10^{-4}
		$\mathcal{S}_{\mathrm{p}}$	6.9×10^{-2}	4.5×10^{-3}	1.1×10^{-14}
	$\frac{L_w}{8}$	$\mathcal{S}$	8.0×10^{-3}	1.4×10^{-2}	1.3×10^{-2}
		$\mathcal{S}_{\mathrm{p}}$	8.0×10^{-3}	1.4×10^{-2}	1.3×10^{-2}



(a) 離散 STFT 後のパワー　(b) FUSTFT 後のパワー　(c) 離散 STFT 後の位相　(d) FUSTFT 後の位相

図 1: $\xi = \frac{L_w}{2} = 256$ のときの離散 STFT とタイプ II の FUSTFT のパワースペクトログラム及び位相スペクトログラム

男性の声」を用いた [15]．$\boldsymbol{x}$ をタイプ II の FUSTFT で複素スペクトログラム $\boldsymbol{X}$ に変換し[18]，これに分散 10^{-6} の複素ガウス雑音を加え $\widehat{\boldsymbol{X}}$ とした．$\boldsymbol{X}$ と $\widehat{\boldsymbol{X}}$ からの逆変換結果 $\hat{\boldsymbol{x}}$ の性能をそれぞれ表 1 と表 2 に示す[19]．表から，2 種類の逆変換がどちらも正しく機能していることが確認できる．特に $\xi = \frac{L_w}{2}$ とした場合には，周期条件を仮定した逆変換を用いれば，複素スペクトログラムが値域 $\mathcal{R}_{\mathcal{S}}$ から外れても，先頭及び最終フレーム以外の誤差をほぼ 0 にできる．

最後に，$\xi = \frac{L_w}{2}$ として式 (15) の離散 STFT とタイプ II の FUSTFT のスペクトログラム[20]を比べる．音信号 $\boldsymbol{x}$ には約 0.2 秒の「クリック音」を用いた [15]．各スペクトログラムを図 1 に示す[21]．滑らかな窓関数を乗算することで信号のエネルギーが周波数方向に広がるため，図 1(b) のように周波数方向を間引いても信号の特徴が保持される．

6　おわりに

本研究では，離散 STFT の半分の周波数成分のみを計算する FUSTFT とその 2 種類の逆変換を提案した．FUSTFT を用いることで，スペクトログラムの行列表現が過度に冗長になることを防ぐことができるため，カイザー窓や打ち切りガウス窓などの大きな窓長 L_w を持つ窓関数が使いやすくなる．$\xi = \frac{L_w}{2}$ とすれば先頭及び最終フレーム以外は振幅と位相を自由に設定できるため，スペクトログラムを用いた信号処理技術の更なる発展も期待される．

[18] 正規化ハニング窓 $w[\tau] := \frac{1}{2\sqrt{L_w}}(1-\cos(\frac{2}{L_w}(\tau+\frac{1}{2})\pi))$ を用いた．

[19] $\|\cdot\|_{\mathrm{F}}^{\mathrm{int}}$ は最初の $\lceil \frac{L_w-\xi}{\xi} \rceil$ フレーム及び最後の $\lceil \lceil \frac{L_x+L_w-\xi}{\xi} \rceil - \frac{L_x}{\xi} \rceil$ フレームの成分を全て 0 とした場合のフロベニウスノルムを意味する．

[20] $L_w = 512$ でサイン窓 $w[\tau] := \frac{1}{\sqrt{L_w}}\sin(\frac{1}{L_w}(\tau+\frac{1}{2})\pi)$ を用いた．

[21] 各スペクトログラムの周波数の数が半分であることに注意されたい．例えば，$\mathrm{FUSTFT_{II}}(\boldsymbol{x})[k,l]$ と $\mathrm{FUSTFT_{II}}(\boldsymbol{x})[\frac{L_w}{2}-k-1,l]$ は複素共役の関係になるため，$k = 0, 1, \ldots, \frac{L_w}{4}-1$ のみで十分となる．この性質のため，FUSTFT は実数値信号に対してもほぼ非冗長な変換となる．また図 1(c), (d) では，振幅が 10^{-4} 未満の複素数の位相を 0 としている．

参考文献

[1] J. B. Allen, “Short term spectral analysis, synthesis, and modification by discrete Fourier transform,” *IEEE Trans. Acoust. Speech Signal Process.*, vol. 25, no. 3, pp. 235–238, 1977.

[2] H. G. Feichtinger and T. Strohmer, Eds., *Gabor Analysis and Algorithms: Theory and Applications*. Secaucus, NJ: Birkhäuser, 1997.

[3] K. Gröchenig, *Foundations of Time-Frequency Analysis*. Secaucus, NJ: Birkhäuser, 2001.

[4] B. Yang, “A study of inverse short-time Fourier transform,” in *Proc. ICASSP*, Las Vegas, NV, 2008, pp. 3541–3544.

[5] K. Yatabe, Y. Masuyama, T. Kusano, and Y. Oikawa, “Representation of complex spectrogram via phase conversion,” *Acoust. Sci. & Tech.*, vol. 40, no. 3, pp. 170–177, 2019.

[6] 若林 佑幸, 小野 順貴, “周波数を冗長化した STFT による位相復元の音声強調への利用,” 日本音響学会 2019 年秋季研究発表会, 草津, 2019, pp. 247–248.

[7] S. Moreno-Picot, F. J. Ferri, M. Arevalillo-Herráez, and W. Díaz-Villanueva, “Efficient analysis and synthesis using a new factorization of the Gabor frame matrix,” *IEEE Trans. Signal Process.*, vol. 66, no. 17, pp. 4564–4573, 2018.

[8] H. S. Malvar, “Lapped transforms for efficient transform/sub-band coding,” *IEEE Trans. Acoust. Speech Signal Process.*, vol. 38, no. 6, pp. 969–978, 1990.

[9] H. S. Malvar, “Fast algorithm for the modulated complex lapped transform,” *IEEE Signal Process. Lett.*, vol. 10, no. 1, pp. 8–10, 2003.

[10] L. Stanković, “On the STFT inversion redundancy,” *IEEE Trans. Circuits Syst. II: Express Briefs*, vol. 63, no. 3, pp. 284–288, 2016.

[11] L. H. Thomas, “Elliptic problems in linear differential equations over a network,” *Watson Sci. Comput. Lab. Report*, Columbia Univ., 1949.

[12] S. Noschese, L. Pasquini, and L. Reichel, “Tridiagonal Toeplitz matrices: properties and novel applications,” *Numer. Linear Algebra Appl.*, vol. 20, no. 2, pp. 302–326, 2013.

[13] S. A. Martucci, “Symmetric convolution and the discrete sine and cosine transforms,” *IEEE Trans. Signal Process.*, vol. 42, no. 5, pp. 1038–1051, 1994.

[14] M. M. Chawla and R. R. Khazal, “A parallel elimination method for “periodic” tridiagonal systems,” *Int. J. Comput. Math.*, vol. 79, no. 4, pp. 473–484, 2002.

[15] MathWorks, “Modified discrete cosine transform: MATLAB mdct.” [Online] https://www.mathworks.com/help/audio/ref/mdct.html