# An Efficient Forecasting Approach to Reduce Boundary Effects in Real-Time Time-Frequency Analysis

Adrien Meynard, Hau-Tieng Wu

# An Efficient Forecasting Approach to Reduce Boundary Effects in Real-Time Time-Frequency Analysis

Adrien Meynard, Hau-Tieng Wu

*Abstract*—Time-frequency (TF) representations of time series are intrinsically subject to the boundary effects. As a result, the structures of signals that are highlighted by the representations are garbled when approaching the boundaries of the TF domain. In this paper, for the purpose of real-time TF information acquisition of nonstationary oscillatory time series, we propose a numerically efficient approach for the reduction of such boundary effects. The solution relies on an extension of the analyzed signal obtained by a forecasting technique. In the case of the study of a class of locally oscillating signals, we provide a theoretical guarantee of the performance of our approach. Following a numerical verification of the algorithmic performance of our approach, we validate it by implementing it on biomedical signals.

*Index Terms*—Boundary effects, time-frequency, forecasting, nonstationarity

## I. INTRODUCTION

IN any digital acquisition system, the study and the interpretation of the measured signals generally require an analysis tool, which enables researchers to point out the useful characteristics of the signal. The need for signal analysis arises from various signals, ranging from audio [1], [2], mechanical [3], or biomedical signals [4]. For instance, biomedical signals, such as photoplethysmogram (PPG), contain several characteristics, including respiratory rate or blood pressure, that cannot be interpreted from its run-sequence plot in the time domain. An analysis tool would make possible the extraction of these useful characteristics.

Usually, the measured signals exhibit nonstationary behavior, and the observed quantities might be interfered with by transient phenomena that can vary rapidly and irregularly. In this paper, we focus on oscillatory time series. These signals might, for example, oscillate fast with large amplitudes at one moment, and then oscillate slowly with small amplitudes at the next moment. In order to adapt the analysis to nonstationarities, local spectral analysis is generally performed [5], [6]. The

A. Meynard is with the Department of Mathematics, Duke University, Durham, NC, 27708 USA.

H.-T. Wu is with the Department of Mathematics and Department of Statistical Science, Duke University, Durham, NC, 27708 USA; Mathematics Division, National Center for Theoretical Sciences, Taipei, Taiwan.

A. Meynard is the corresponding author (e-mail: adrien.meynard@duke.edu).

short-time Fourier transform [7] (STFT), a typical tool built for this purpose, enables the determination of the local frequency content of a nonstationary signal.

Windowing is a common method for performing local analysis. Among many others, STFT [8], continuous wavelet transform (CWT) [9], synchrosqueezing transform (SST) [10], and reassignment [11] (RS) are representations that fall back on the use of an analysis window. Let $x : I \to \mathbb{R}$ denote the observed signal, where $I$ denotes the finite interval where the signal is measured. Let $g_s : \mathbb{R} \to \mathbb{R}$ denote the analysis window, where $s$ is a shape parameter. The support of $g_s$ is localized around the origin and is small with respect to $|I|$. The translation operator is $T_\tau$ defined as:

$$T_\tau f = f(t - \tau) , \quad \forall f : \mathbb{R} \to \mathbb{R} .$$

Then, the local analysis of $x$ around the instant $\tau \in I$ relies on the evaluation of the following inner product:

$$V_x(s, \tau) = \langle x, T_\tau g_s \rangle_I . \tag{1}$$

A major shortcoming of this technique occurs when analyzing the signal $x$ *near the boundaries* of the interval $I$. Clearly, at these points, half of the information is missing. Consequently, the results of the inner product (1) are distorted. This phenomenon is usually understood as the *boundary effect*. The result of the SST of a PPG is shown in the lower-left corner of Fig. 1 (see Section IV-D3 for a comprehensive description). The distortion resulting from boundary effects is clearly visible on the right side of this representation—this area is highlighted by the red dashed rectangle. Indeed, while in the major left part of the image, clear lines stand out, they become blurred as they approach the right boundary of the image. Therefore, estimations of signal characteristics, like instantaneous frequencies [12] or amplitudes, from this TF representation appear to be imprecisely determinable (or even likely to fail) in the vicinity of the boundaries. Moreover, this boundary effect would unavoidably limit applying these window-based TF analysis tools for real-time analysis purposes. It is thus desirable to have a solution to eliminating the boundary effects.

Attempts to minimize the boundary effects generally consists in softening the discontinuity on signal edges. Usually, the approaches can be classified into two main
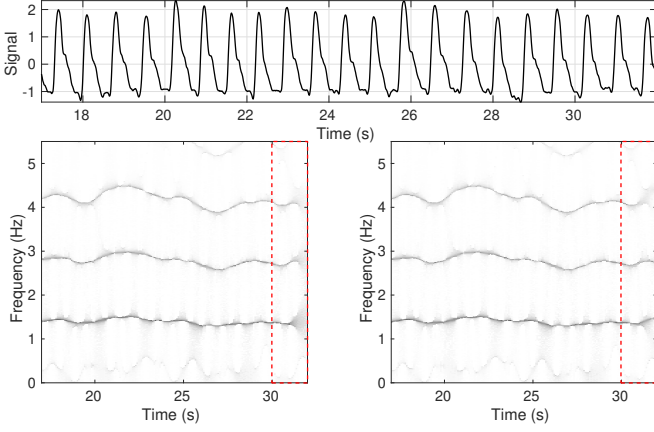
Fig. 1. A segment of PPG signal (top) and the right boundary of a TF representation determined by the SST without extension (bottom left), and the right boundary of a TF representation determined by the SST with the proposed boundary effect reduction algorithm by forecasting (bottom right). The window length for the SST is 12 seconds. Videos of these real-time SSTs are available online at https://github.com/AdMeynard/BoundaryEffectsReduction.

classes, choosing a proper analysis window and extending/forecasting the signal. In the first class, judicious choice of analysis window whose support does not interfere with the boundary points can minimize the occurrence of aberrant patterns near the boundaries of the TF plane, for instance, [13], [14]. Due to the specific relationship between the chosen analysis windows and the TF analysis tool, these techniques do not make it possible to reduce the boundary effects of all TF representations. Another natural idea, the second class, consists of carrying out a preliminary step of extending, or forecasting, the signal beyond its boundaries. Due to its flexibility, various forecasting schemes have been proposed. For example, there exist simple extension schemes that do not take into account the dynamical behavior of the signal, such as zero-padding, periodic extension, symmetric extension [15], [16], or polynomial extrapolation [17]. There exist extension schemes based on physically relevant dynamical models, such as the Extended Dynamic Mode Decomposition [18] (EDMD) and the Gaussian process regression [19], [20] (GPR). In speech processing, dynamic mode predictors have also been proposed [21] to forecast signals falling into the so-called *source–filter* model. This gradient-based technique relies on the shadowing approach proposed in [22]. There also exist extension schemes based on stochastic models, such as the Trigonometric, Box-Cox transformation, ARMA errors, Trend and Seasonal components (TBATS) algorithm [23], the dynamic linear models [24]. In the physically relevant dynamical models, the oscillation or trend, are usually modeled as the mean of a stochastic process, while in the stochastic models, the oscillation or trend are modeled by the covariance structure of the stochastic process. It is also possible to consider polynomial regression [25] or modeling the mean by splines [26], kernel functions [27], and wavelets [28],

or nonparametric regression [25], to estimate the mean function before forecasting the signal. Neural network models, such as the long short-term memory model [29], is another approach. The above list is far from exhaustive, and we refer readers with interest to [30] for a friendly monograph on the general forecasting topic. While the model-based extension scheme gives better-extended signals than those simple extension without dynamical models, they generally have a great computational cost.

In this paper, we propose a fast extension algorithm based on a simple dynamical model that optimizes the trade-off between the extension quality and the computational cost, so that the real-time analysis can be achieved. The algorithm is composed of two steps.

1) *Extend the signal by forecasting it.* The aim is to use a simple dynamic model to predict the values taken by the measured signal outside the measurement interval. Then, once this operation is done, we have access to an extended signal defined on a larger interval $I_\Delta$, where $\Delta$ denotes the size of the extension on both boundaries of $I$.
2) *Run the local analysis tool on the extended signal.* Assuming that the support of the analysis window is smaller than $2\Delta$, the local analysis near the boundary of $I$ is now possible without lack of information thanks to knowledge brought by the extension.

Thus, assuming that the quality of the extension step is sufficient, the analysis results obtained that way will be less sensitive to the boundary effects than the result of the analysis tool applied directly to the nonextended signal. We claim and prove that forecasting oscillatory signals based on the simple dynamical model combined with the simple least square approach is sufficient for reducing the boundary effects for the TF analysis, or other kernel-based analysis. To our knowledge, such theoretical analysis does not exist in the literature. See the bottom of Fig. 1—in particular, the regions indicated by the dashed boxes—for a snapshot of the result. The main benefit of this simple approach is a numerically efficient solution with a theoretical guarantee for real-time analysis purposes.

The paper is organized in the following way. In section II, we provide an extension method based on a linear dynamic model. We derive the corresponding algorithm for boundary effects reduction. In Section III, we show that the dynamic model we consider is sufficient to extend signals taking the form of sums of sine waves. An evaluation of the theoretical performance of our algorithm on this class of signals is given in Section III. In Section IV, we compare our extension method with more sophisticated methods such as EDMD, GPR, or TBATS. We show that our algorithm gives fast results of reasonable quality. Finally, we evaluate the performance of our boundary effects reduction algorithm on biomedical signals, such as respiratory signals, and compare it to the theoretical results.

## II. Algorithm

As explained above, the algorithm for the reduction of boundary effects on TF representations relies on extending the signal by forecasting it before applying the TF analysis.

We start with the notation. Let $x : \mathbb{R} \to \mathbb{R}$ denote a continuous-time signal. In this work, we consider a finite-length discretization of that one. Thus, the sampled signal $\mathbf{x}$, whose length is denoted by $N$, is such that

$$\mathbf{x}[n] = x\left(\frac{n}{f_\mathsf{s}}\right) , \quad \forall n \in \{0, \dots, N-1\} ,$$

where $f_\mathsf{s}$ denotes the sampling frequency. Let $M$ and $K$ be two positive integers such that $M < K$ and $K + M \leq N$. Then, for all $k \in \{0, \dots, K-1\}$, we extract from $\mathbf{x} \in \mathbb{R}^N$ the subsignal $\mathbf{x}_k \in \mathbb{R}^M$ given by:

$$\mathbf{x}_k = \begin{pmatrix} \mathbf{x}[N-K-M+k] \\ \vdots \\ \mathbf{x}[N-K+k-1] \end{pmatrix} . \tag{2}$$

These subsignals are gathered into the matrix $\mathbf{X} \in \mathbb{R}^{M \times K}$ such that:

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_0 & \cdots & \mathbf{x}_{K-1} \end{pmatrix} .$$

Notice that these subsignals are overlapping each other. Indeed, $\mathbf{x}_{k+1}$ is a shifting of $\mathbf{x}_k$ from one sample. We also consider the matrix $\mathbf{Y} \in \mathbb{R}^{M \times K}$ given by:

$$\mathbf{Y} = \begin{pmatrix} \mathbf{x}_1 & \cdots & \mathbf{x}_K \end{pmatrix} .$$

The boundary effect reduction algorithm is based on manipulating $\mathbf{X}$ and $\mathbf{Y}$.

The pseudo-code of the proposed real-time algorithm to reduce boundary effects on windowing-based TF representations is shown in Algorithm 1. We coined the algorithm BoundEffRed. Below, we detail the algorithm, particularly the signal extension SigExt in Algorithm 2.

---

**Algorithm 1** Tackling boundary effects of a TF representation in real-time. $\mathbf{F_x} = \mathsf{BoundEffRed}(\mathbf{x}, M, K, L, \mathscr{F})$

---

**Inputs**: $\mathbf{x}_N$, $M$, $K$, $N_0$, $L$, $\mathscr{F}$

**while** $N$ increases **do**
  **Real-time input**: $\mathbf{x}_N$

  **Forecasting step.**
    • Signal extension: $\tilde{\mathbf{x}}_{N+L} = \mathsf{SigExt}(\mathbf{x}_N)$.
  **Representation estimation step.**
    • Extended representation evaluation: $\mathscr{F}(\tilde{\mathbf{x}}_{N+L})$.
    • Restriction of $\mathscr{F}(\tilde{\mathbf{x}}_{N+L})$ to the current time interval (see (9)) to obtain $\mathbf{F}_\mathbf{x}^{(N)} = \mathscr{F}^{\text{ext}}(\mathbf{x}_N)$.

  **Real-time output**: Signal representation $\mathbf{F}_\mathbf{x}^{(N)}$
**end while**

---

### A. Step 1: Extension by Forecasting

*a) Dynamical Model:* Establishing a dynamical model means determining the relation linking $\mathbf{Y}$ to $\mathbf{X}$; that is, finding a function $f$ so that

$$\mathbf{Y} = f(\mathbf{X}) . \tag{3}$$

In a general framework, forecasting means estimating the function $f$ from the observed values taken by the signal, in order to predict its future values. For instance, the dynamic mode decomposition [31], [18] or other more complicated models [20], [23], [24], [29], allow this by setting additional constraints on the behavior of $f$. We will see, in Section III, that considering such a complex dynamic model is not necessary for the study of the oscillatory signals of interest to us. That is why we consider here a naive dynamical model, assuming that we have the following relation:

$$\mathbf{Y} = \mathbf{A}\mathbf{X} , \tag{4}$$

where $\mathbf{A} \in \mathbb{R}^{M \times M}$. In other words, we adopt a classical strategy in the study of dynamical systems, the *linearization of a nonlinear system*, when the system is sufficiently regular. Notice that this linearized dynamical model can be written equivalently according to the subsignals $\mathbf{x}_k$, as:

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k, \, , \forall k \in \{0, \dots, K-1\} . \tag{5}$$

*b) Forecasting:* Determining the forecasting amounts to estimating the unknown matrix $\mathbf{A}$. Indeed, let $\tilde{\mathbf{A}}$ denotes the estimate of $\mathbf{A}$. We then obtain the forecasting of $\mathbf{x}_{k+\ell}$ by recursively applying the linear relation (4):

$$\tilde{\mathbf{x}}_{K+\ell} = \underbrace{\tilde{\mathbf{A}}\tilde{\mathbf{A}}\cdots\tilde{\mathbf{A}}}_{\ell \text{ times}} \mathbf{x}_K = \tilde{\mathbf{A}}^\ell \mathbf{x}_K .$$

Let $\mathbf{e}_M$ denote the unit vector of length $M$ given by $\mathbf{e}_M = \begin{pmatrix} 0 & \cdots & 0 & 1 \end{pmatrix}^T$. Consequently, given definition (2), the forecasting of the signal at time $\frac{N-1+\ell}{f_\mathsf{s}}$ is provided by

$$\tilde{\mathbf{x}}[N-1+\ell] = \mathbf{e}_M^T \tilde{\mathbf{x}}_{K+\ell} = \boldsymbol{\alpha}^{(\ell)} \mathbf{x}_K , \tag{6}$$

where $\boldsymbol{\alpha}^{(\ell)}$ denotes the last row of $\tilde{\mathbf{A}}^\ell$, i.e., $\boldsymbol{\alpha}^{(\ell)} = \mathbf{e}_M^T \tilde{\mathbf{A}}^\ell$.

*c) Model Estimation:* To estimate the matrix $\mathbf{A}$, we consider the simple but numerically efficient least square estimator. That is, we solve the following problem:

$$\tilde{\mathbf{A}} = \arg\min_{\boldsymbol{\alpha}} \mathcal{L}(\mathbf{A}) , \tag{7}$$

where the loss function $\mathcal{L}$ is given by:

$$\mathcal{L}(\mathbf{A}) = \|\mathbf{Y} - \mathbf{A}\mathbf{X}\|^2 = \sum_{k=0}^{K-1} \|\mathbf{x}_{k+1} - \mathbf{A}\mathbf{x}_k\|^2.$$

Therefore, solving the problem (7), i.e., $\nabla \mathcal{L}(\tilde{\mathbf{A}}) = \mathbf{0}$, gives the following estimate $\tilde{\mathbf{A}}$ of the dynamical model matrix $\mathbf{A}$:

$$\tilde{\mathbf{A}} = \mathbf{Y}\mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1} . \tag{8}$$

**Remark 1.** *This expression clearly shows that the matrix $\tilde{\mathbf{A}}$ takes the following form:*

$$\tilde{\mathbf{A}} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 \\ \alpha_1 & \cdots & \cdots & \cdots & \alpha_M \end{pmatrix} .$$

*Then, except for the row vector $\boldsymbol{\alpha} = (\alpha_1 \cdots \alpha_M)$, the matrix $\mathbf{A}$ is fully determined by the dynamical model.*

*d) Signal Extension:* Since we are building a real-time algorithm, we consider that only the right "side" of the signal has to be extended, the left "side" being fixed since it only concerns the past values of the signal. We therefore construct the extended signal $\tilde{\mathbf{x}} \in \mathbb{R}^{N+L}$ concatenating the observed signal $\mathbf{x}$, and the forward forecasting $\tilde{\mathbf{x}}_{\text{fw}} \in \mathbb{R}^L$. We summarize the extension step in Algorithm 2.

---

**Algorithm 2** Signal extension. $\tilde{\mathbf{x}} = \mathsf{SigExt}(\mathbf{x}, M, K, L)$

**Inputs**: $\mathbf{x}_N$, $M$, $K$, $L$

**Forward forecasting.**
- Estimation of the matrix $\tilde{\mathbf{A}}$ via equation (8).
- Forecasting $\tilde{\mathbf{x}}_{\text{fw}} \in \mathbb{R}^L$ obtained applying equation (6) with $\ell \in \{1, \ldots, L\}$.

**Output**: Extended signal $\tilde{\mathbf{x}}_{N+L} = \begin{pmatrix} \mathbf{x}_N & \tilde{\mathbf{x}}_{\text{fw}} \end{pmatrix}^T$.

---

*B. Step 2: Extended Time-Frequency Representation*

Let $\mathscr{F} : \mathbb{R}^N \to \mathbb{C}^{F \times N}$ generically denotes the TF representation of interest to us, which could be, for instance, STFT, CWT, SST, or RS. Here, $F$ typically denotes the size of the discretization along the frequency axis. Due to the boundary effects, the representation $\mathscr{F}(\mathbf{x}_N)$ shows undesired patterns near its edges. To alleviate the boundary effects, we apply the representation to the estimated extended signal $\tilde{\mathbf{x}}$. This strategy moves the boundary effects out of the time interval $I = [0, \frac{N-1}{f_s}]$. Finally, the boundary-effects insensitive representation $\mathscr{F}^{\text{ext}} : \mathbb{R}^N \to \mathbb{C}^{F \times N}$ of $\mathbf{x}_N$ is given for all $\nu \in \{0, \cdots, F-1\}$, $n \in \{0, \cdots, N-1\}$ by:

$$\mathscr{F}^{\text{ext}}(\mathbf{x}_N)[\nu, n] = \mathscr{F}(\tilde{\mathbf{x}}_{N+L})[\nu, n] . \quad (9)$$

This amounts to restricting the representation $\mathscr{F}(\tilde{\mathbf{x}}_{N+L})$ to the current measurement interval of $\mathbf{x}_N$. For the sake of simplicity, we denote the *restriction operator* by $\mathcal{R}$, where $\mathcal{R} : \mathbb{C}^{F \times (N+L)} \to \mathbb{C}^{F \times N}$. Consequently, we have:

$$\mathscr{F}^{\text{ext}}(\mathbf{x}_N) = \mathcal{R}\left(\mathscr{F}(\tilde{\mathbf{x}}_{N+L})\right) . \quad (10)$$

We call $\mathscr{F}^{\text{ext}}$ the *boundary-free TF representation*.

$\mathbf{F}_{\mathbf{x}}^{(N)} \in \mathbb{C}^{F \times N}$ is the estimation of the boundary-free TF representation at iteration $N$ in Algorithm 1. For numerical purposes, and to make the real-time implementation achievable, we do not perform a full re-estimation of

$\mathscr{F}^{\text{ext}}$ at iteration $N+1$. Instead, the additional knowledge provided by $\mathbf{x}[N+1]$ only influences the values of the last $L_{\text{win}}$ columns of $\mathscr{F}^{\text{ext}}(\mathbf{x}_{N+1})$, where $L_{\text{win}}$ denotes the window half-length used by the TF representation. Thus, $\mathbf{F}_{\mathbf{x}}^{(N+1)}$ is obtained by the concatenation of the first $N - L_{\text{win}} + 1$ columns of $\mathbf{F}_{\mathbf{x}}^{(N)}$ with the last $L_{\text{win}}$ columns of $\mathscr{F}^{\text{ext}}(\mathbf{x}_{N+1})$.

### III. THEORETICAL PERFORMANCE

*A. Signal Model*

We model the meaningful part of the observed signal as a deterministic multicomponent harmonic signal; that is, a sum of sine waves:

$$\mathbf{z}[n] = \sum_{j=1}^{J} \Omega_j \cos\left(2\pi f_j \frac{n}{f_s} + \varphi_j\right) , \quad (11)$$

where $J$ denotes the number of components, $\Omega_j > 0$ the amplitude of the $j$th component, $f_j$ its frequency, and $\varphi_j \in [0, 2\pi)$ its initial phase. For the sake of simplicity, we make an additional assumption on the frequencies of each component. We assume that for all $j \in \{1, \ldots, J\}$:

$$\exists\, p_j, p_j' \in \mathbb{N}^* : \quad f_j = \frac{p_j}{M} f_s = \frac{p_j'}{K} f_s . \quad (12)$$

In addition, the observed signal is assumed to be corrupted by an additive Gaussian white noise. Therefore, the measured discrete signal $\mathbf{x}$ is written as:

$$\mathbf{x} = \mathbf{z} + \sigma \mathbf{w} , \quad (13)$$

where $\mathbf{z}$ follows model (11), $\mathbf{w}$ is a Gaussian white noise, whose variance is normalized to one, and $\sigma > 0$. Clearly, $\sigma^2$ denotes the variance of the additive noise $\sigma \mathbf{w}$.

*B. Forecasting Error*

On the forecasting interval, we decompose the estimated signal $\tilde{\mathbf{x}}$ as follows:

$$\tilde{\mathbf{x}}[n] = \mathbf{z}[n] + \boldsymbol{\epsilon}[n] , \quad (14)$$

where $\boldsymbol{\epsilon}$ is the forecasting error. When $n \in I = \{0, \ldots, N-1\}$, this error contains only the measurement noise, that is $\boldsymbol{\epsilon}[n] = \sigma \mathbf{w}[n]$. Outside the interval $I$, the importance of the forecasting error $\boldsymbol{\epsilon}$ is also affected by the loss of information resulting from the linearization of the dynamical model we consider in (4). To evaluate the actual behavior of the forward forecasting error $\boldsymbol{\epsilon}[n]$ when $n \geq N$, we determine its first two moments.

1) The mean, or estimation bias, is such that:

$$\boldsymbol{\mu}[n] \triangleq \mathbb{E}\{\boldsymbol{\epsilon}[n]\} = \mathbb{E}\{\tilde{\mathbf{x}}[n]\} - \mathbf{z}[n] . \quad (15)$$

Given the forecasting strategy, we have $\boldsymbol{\mu}[n] = 0$ when $n \in I$ and

$$\boldsymbol{\mu}[n] = \mathbb{E}\{\boldsymbol{\alpha}^{(n-N+1)}\}\mathbf{z}_K + \sigma\mathbb{E}\{\boldsymbol{\alpha}^{(n-N+1)}\mathbf{w}_K\} - \mathbf{z}[n] \quad (16)$$

when $n \geq N$.

2) The covariance is given by:

$$\gamma[n, n'] \triangleq \mathbb{E}\{(\boldsymbol{\epsilon}[n] - \boldsymbol{\mu}[n]) (\boldsymbol{\epsilon}[n'] - \boldsymbol{\mu}[n'])\}$$
$$= \mathbb{E}\{\tilde{\mathbf{x}}[n]\tilde{\mathbf{x}}[n']\} - \mathbf{z}[n]\mathbf{z}[n'] - \boldsymbol{\mu}[n]\mathbf{z}[n']$$
$$- \boldsymbol{\mu}[n']\mathbf{z}[n] - \boldsymbol{\mu}[n]\boldsymbol{\mu}[n'] .$$

Thus by definition of the noise, we have $\gamma[n, n'] = \sigma^2\delta_{n,n'}$ when $(n, n') \in I^2$. When $n \geq N$, let us denote $\ell = n - N + 1$. Then, we have two cases.

(i) If $n' \in I$:

$$\gamma[n, n'] = \sigma\mathbb{E}\{\mathbf{w}[n']\boldsymbol{\alpha}^{(\ell)}\}\mathbf{z}_K + \sigma^2\mathbb{E}\{\mathbf{w}[n']\boldsymbol{\alpha}^{(\ell)}\mathbf{w}_K\} .$$
$$(17)$$

(ii) If $n' = N - 1 + \lambda \geq N$:

$$\gamma[n, n'] = \mathbf{z}_K^T\mathbb{E}\left\{\boldsymbol{\alpha}^{(\ell)T}\boldsymbol{\alpha}^{(\lambda)}\right\}\mathbf{z}_K + \sigma\mathbb{E}\{\boldsymbol{\alpha}^{(\ell)}\mathbf{w}_K\boldsymbol{\alpha}^{(\lambda)}\}\mathbf{z}_K$$
$$+ \sigma\mathbb{E}\{\boldsymbol{\alpha}^{(\lambda)}\mathbf{w}_K\boldsymbol{\alpha}^{(\ell)}\}\mathbf{z}_K + \sigma^2\mathbb{E}\{\boldsymbol{\alpha}^{(\ell)}\mathbf{w}_K\boldsymbol{\alpha}^{(\lambda)}\mathbf{w}_K\}$$
$$- \mathbf{z}[n]\mathbf{z}[n'] - \mathbf{z}[n]\boldsymbol{\mu}[n'] - \mathbf{z}[n]\boldsymbol{\mu}[n'] - \boldsymbol{\mu}[n]\boldsymbol{\mu}[n'].$$
$$(18)$$

Besides, we recall that $\gamma[n, n'] = \gamma[n', n]$.

In Theorem 1, we specify the asymptotic behavior of the forecasting bias and covariance when the dataset size $K$ is great.

**Theorem 1.** *Let $\mathbf{x} \in \mathbb{R}^N$ be a discrete-time random signal following model* (13). *Let $\tilde{\mathbf{x}}$ denotes its forecasting, obtained using the extension Algorithm 2. Let $n \geq N$ be a sample index. Then, the first-order moment of the forecasting error $\epsilon[n]$ in* (14) *defined in* (15) *satisfies*

$$|\boldsymbol{\mu}[n]| \leq a_0^{(n)}\sigma^2 + \frac{1}{K}\left(\frac{a_1^{(n)}}{\sigma^2} + a_2^{(n)}\right) + o\left(\frac{1}{K}\right) \quad (19)$$

*as $K \to \infty$, where $a_0^{(n)}$, $a_1^{(n)}$, and $a_2^{(n)}$ are positive quantities, independent of $K$ and $\sigma$. Its second-order moment $\gamma[n, n']$ satisfies the following approximation equations:*

*(i) if $n' \in I = \{0, \ldots, N-1\}$:*

$$|\gamma[n, n']| \leq b_0^{(n,n')}\sigma^2 + \frac{1}{K}\left(b_1^{(n,n')} + b_2^{(n,n')}\sigma^2\right) + o\left(\frac{1}{K}\right)$$
$$(20)$$

*as $K \to \infty$, where $b_0^{(n,n')}$, $b_1^{(n,n')}$, and $b_2^{(n,n')}$ are positive quantities, independent of $K$ and $\sigma$;*

*(ii) if $n' \geq N$:*

$$|\gamma[n, n']| \leq c_0^{(n,n')}\sigma^2$$
$$+ \frac{1}{K}\left(\frac{c_1^{(n,n')}}{\sigma^2} + c_2^{(n,n')} + c_3^{(n,n')}\sigma^2\right) + o\left(\frac{1}{K}\right)$$
$$(21)$$

*as $K \to \infty$, where $c_0^{(n,n')}$, $c_1^{(n,n')}$, $c_2^{(n,n')}$ and $c_3^{(n,n')}$ are positive quantities, independent of $K$ and $\sigma$.*

*Proof.* See the Supplementary Materials. The proof is mainly based on Taylor expansions of the forecasting error. Nonoptimal values of $a_0^{(n)}$, $a_1^{(n)}$, ..., $b_0^{(n,n')}$, ..., $c_3^{(n,n')}$ are also provided in the proof. ∎

The forecasting bias and covariance asymptotically depend linearly on the ratio $\frac{1}{K}$. This shows the need to use a sufficiently large dataset to obtain an accurate forecast. Ideally when $K \to \infty$, the forecasting error would behave like the measurement noise $\sigma\mathbf{w}$, i.e., a zero-mean white noise whose variance is of the order of $\sigma^2$. However, Theorem 1 shows that the estimator may remain asymptotically biased (the first term in the bound of equation (19) being independent of $K$). Concerning the covariance of the forecasting error, the expected asymptotic behavior is verified. Indeed, when $K \to \infty$, the variance of the forecasting estimator increases at most linearly with the noise variance $\sigma^2$, since is bounded by $b_0^{(n,n')}\sigma^2$.

Assume now that $K$ is sufficiently great and fixed. The noise influences the quality of the forecasting estimator in two ways. On the one hand, when the noise variance $\sigma^2$ is great, the bias and the variance of the estimator potentially increase linearly with $\sigma^2$. This reflects the classical behavior of an estimator of a signal degraded by additive noise. On the other hand, due to terms in $1/\sigma^2$, when the noise variance $\sigma^2$ is small, the bias and the variance of the forecasting estimator are no longer controlled via equations (19) and (21). This illustrates the necessary presence of noise to obtain an accurate prediction. Indeed, if $\sigma$ is too small, the matrix $\mathbf{XX}^T$ is ill-conditioned and its inversion in (8) is sensitive. The forecasting is then of poor quality. Noise therefore appears as a regularization term to ensure the invertibility of $\mathbf{XX}^T$. This seemingly counterintuitive fact indicates the challenge we encounter to do forecasting with the dynamic model (3), the degeneracy. Note that from the lag map perspective, $\mathbf{X}$ and $\mathbf{Y}$ consist of points located on a low dimensional manifold. Thus, locally the ranks of $\mathbf{X}$ and $\mathbf{Y}$ are low, which leads to the degeneracy. However, when noise exists, it naturally fills in deficient ranks, and stabilizes the algorithm.

The dependencies of the forecasting quality on the subsignals lengths $M$ and the forecasting index $\ell = n - N + 1$ are hidden in the expression of the parameters $a_0^{(n)}$, $a_1^{(n)}$, ..., $b_0^{(n,n')}$, ..., $c_3^{(n,n')}$. The numerical results, discussed in Section IV-C1, allow us to better evaluate these dependencies.

*C. Adaptive Harmonic Model*

One can extend the previous result to the case where the instantaneous frequencies and amplitudes of the components of the deterministic part of the observed signal are *slowly varying*. It is an *AM-FM model* that, in its is continuous-time version, takes the following form

$$z(t) = \sum_{j=1}^{J} a_j(t)\cos(2\pi\phi_j(t)), \quad (22)$$

where $a_j(t)$ and $\phi_j'(t)$ describe how large and fast the signal oscillates at time $t$. Clearly, (11) is a special case satisfying the AM-FM model when the $a_j$ and $\phi_j'$ are

both constants. In many practical signals, the amplitude and frequency do not change fast. It is thus reasonable to further restrict the regularity and variations of the instantaneous amplitudes and frequencies of these AM-FM functions. When the speeds of variation of the instantaneous amplitudes $a_j$ and frequencies $\phi_j'$ are small, the signal can be "locally" well approximated by a harmonic function in (11); that is, locally at time $t_0$, $z(t)$ can be well approximated by $z_0(t) := \sum_{j=1}^{J} a_j(t_0) \cos(2\pi(\phi_j(t_0) - t_0\phi_j'(t_0) + \phi_j'(t_0)t))$ by the Taylor expansion. An AM-FM function satisfying the slow variation of the instantaneous amplitudes $a_j$ and frequencies $\phi_j'$ is referred to the adaptive harmonic model (AHM) (see [32], [33] for mathematical details). It is thus clear that the forecasting error caused by the SigExt algorithm additionally depends on the speed of variation of the instantaneous amplitudes $a_j$ and frequencies $\phi_j'$. For the forecasting purpose, it is thus clear that when the speed of variation of $a_j$ and $\phi_j'$ are slow $K$ can be chosen large while the signal can be well approximated by (11). Hence, Theorem 1 can still be applied to approximate the error. However, how to determine the optimal $K$ based on the signal depends on the application and is out of the scope of this paper. It will be explored in future work for specific application problems.

The models described in Sections III-A and III-C consider the meaningful part of the signal to be a deterministic component. These models are purposely adapted to signals showing local line spectra. The physiological signals we are interested in, such as respiratory or cardiac signals, have this characteristic (see Section IV-D). Signals with wider local spectra, such as electroencephalography signals, do not fall into this category, and are more faithfully modeled as random signals. Thus, the theoretical justifications proposed above are no longer applicable to guarantee the forecasting quality of SigExt.

### D. Performance of the Boundary Effects Reduction

While we do not provide a generic proof of the boundary effect reduction on *any* TF analysis tools, we discuss a particular case of SST. Since SST is designed to analyze signals satisfying the AHM, as is discussed in Section III-C, Theorem 1 ensures that the forecasting error $\epsilon$ defined in (14) is controlled and bounded in terms of mean and covariance. Recall Theorem 3 in [32], which states that when the additive noise is stationary and may be modulated by a smooth function with a small fluctuation, the SST of the observed signal remains close to the SST of the clean signal, throughout the TF plane. We refer the reader to [32] for a precise quantification of the error made in the TF plane, which depends notably on the covariance of the additive noise and on the speed of variation of the amplitudes and instantaneous frequencies composing the signal. In our case, while the noise of the historical data is stationary, the forecasting error depends on the historical noise, and hence the overall "noise" is nonstationary. However, the dependence only appears in the extended part of the signal. We claim that the proof of Theorem 3 in [32] can be generalized to explain the robustness of SST to the noise, and a systematic proof will be reported in our future theoretical work. We verify experimentally that Algorithm 1 is efficient for a large number of representations in the following Section IV. Therefore, in our case, this means that the boundary effect is strongly reduced since the impact of the forecasting error does SST is limited. An immediate application is that the instantaneous frequencies can now be well estimated continuously up to the edges of the TF plane, and real-time acquisition of the instantaneous frequency and amplitude information is possible.

## IV. NUMERICAL RESULTS

For the reproducibility purpose, the MATLAB code and datasets used to produce the numerical results in this section are available online at https://github.com/AdMeynard/BoundaryEffectsReduction.

### A. Extension Methods for Comparison

To highlight the fact that the linear dynamical model is sufficient to catch most of the dynamical behavior of signals following the AHM, we compare the performance of Algorithm 1 with a simple extension obtained by pointwise symmetrization [15]. We also evaluate the performance of reference forecasting algorithms that could be used for extending such signals. These methods are:

- The EDMD has been developed by Williams *et al.* [18]. The proposed algorithm is a way to obtain an approximation of the so-called Koopman operator of the observed system, which theoretically allows catching dynamic of nonlinear systems [34].
- The GPR [19] is a method relying on a probabilistic dynamical model. That one is based on the Gaussian process structure, and therefore offers more flexibility in the type of dynamic that could be modeled than the linear model (4).
- The TBATS method [23] is based on a classical decomposition of times series into a trend, seasonal and ARMA components, with a specific dynamic for the seasonal component. This model demands the estimation of numerous parameters and, by implication, may require a long computation time.

### B. Evaluation Metrics

The first evaluation metric is quantifying the forecasting quality (i.e. not depending on $\ell$) in the time domain. We evaluate the Experimental Mean Square Error $\mathrm{MSE_{xp}}(\tilde{\mathbf{x}})$ of the forward forecast extended signals, namely,

$$\mathrm{MSE_{xp}}(\tilde{\mathbf{x}}) = \frac{1}{L}\|\tilde{\mathbf{x}} - \mathbf{x}^{\mathrm{ext}}\|^2 \tag{23}$$
$$= \frac{1}{L}\sum_{\ell=1}^{L} \boldsymbol{\mu}_{\mathrm{xp}}[N-1+\ell]^2 + \gamma_{\mathrm{xp}}[N-1+\ell, N-1+\ell].$$

where $\mathbf{x}^{\text{ext}}$ is the *ground-truth extended* signal; that is, $\mathbf{x}^{\text{ext}} := \left( \mathbf{x}[-L] \quad \cdots \quad \mathbf{x}[N-1+L] \right)$. Then, as long as the bias $\boldsymbol{\mu}[N-1+\ell]$ and the variance $\gamma[N-1+\ell, N-1+\ell]$ of the forecasting estimator remain small for all $\ell$, the MSE takes small values either.

The second evaluation metric is evaluating the quality of the boundary effect reduction directly on the TF representation. We compare the obtained TF representation to the optimal TF representation $\mathscr{F}_N^{\text{opt}}(\mathbf{x})$, defined as the restriction of the representation of the ground-truth extended signal $\mathbf{x}^{\text{ext}}$; that is, we have

$$\mathscr{F}^{\text{opt}}(\mathbf{x}_N) = \mathcal{R}\left(\mathscr{F}(\mathbf{x}^{\text{ext}})\right), \tag{24}$$

where $\mathcal{R}$ is defined in (10). To compare different techniques, we use a criterion proposed in [33] that quantifies the optimal transport (OT) distance between a given TF representation and the optimal one. In general, the OT distance is a quantity that measures how different two probability density functions are. Denote a TF representation as $\mathcal{Q}$. For $t$ fixed, we consider the probability density function defined by $p_{\mathcal{Q}}^t(\xi) = |\mathcal{Q}(\xi,t)|^2 / \int_{\mathbb{R}} |\mathcal{Q}(\nu,t)|^2 \mathrm{d}\nu$; that is, normalizing the spectral information at time $t$. At each time $t$, the OT distance $d_t$ between two TF representations $\mathcal{Q}$ and $\mathscr{F}^{\text{opt}}$ at time $t$ is defined by

$$d_t(\mathcal{Q}, \mathscr{F}^{\text{opt}}) = \int_{\mathbb{R}} \left| \tilde{P}_{\mathcal{Q}}^t(\xi) - P_{\mathscr{F}^{\text{opt}}}^t(\xi) \right| \mathrm{d}\xi,$$

where $P_{\mathcal{Q}}^t(\xi) = \int_{-\infty}^{\xi} p_{\mathcal{Q}}^t(\nu) \mathrm{d}\nu$ and $\tilde{P}_{\mathscr{F}^{\text{opt}}}^t(\xi) = \int_{-\infty}^{\xi} \tilde{p}_{\mathscr{F}^{\text{opt}}}^t(\nu) \mathrm{d}\nu$. In other words, the OT distance between $\mathcal{Q}$ and $\mathscr{F}^{\text{opt}}$ at time $t$ is given by the $L^1$ norm of the associated cumulative density function. Note that $d_t(\mathcal{Q}, \mathscr{F}^{\text{opt}})$ quantifies the proximity between the estimated and actual TF representations at time $t$. With the OT distance, we define the *performance index* $D(\mathcal{Q}, \mathscr{F})$ for the reduction of boundary effects of a given TF representation $\mathcal{Q}$ over another TF representation $\mathscr{F}$ by

$$D(\mathcal{Q}, \mathscr{F}) = \frac{\int_I d_t\left(\mathcal{Q}, \mathscr{F}^{\text{opt}}\right) \mathrm{d}t}{\int_I d_t\left(\mathscr{F}, \mathscr{F}^{\text{opt}}\right) \mathrm{d}t}. \tag{25}$$

In other words, $D(\mathcal{Q}, \mathscr{F})$ is the ratio between its averaged OTD to the optimal TF representation $\mathscr{F}^{\text{opt}}$ (see (24)) and the averaged OTD of the original TF representation $\mathscr{F}$ to $\mathscr{F}^{\text{opt}}$. Thus, $D(\mathcal{Q}) < 1$ means a reduction of the boundary effects. Let us evaluate the quality of the boundary effects reduction on biomedical signals.

### C. Simulated Signals

We first evaluate the quality of the forecasting step and compare it to the theoretical results provided by Theorem 1. The level of the forecasting error depends on at least two parameters:

- The noise variance $\sigma^2$.
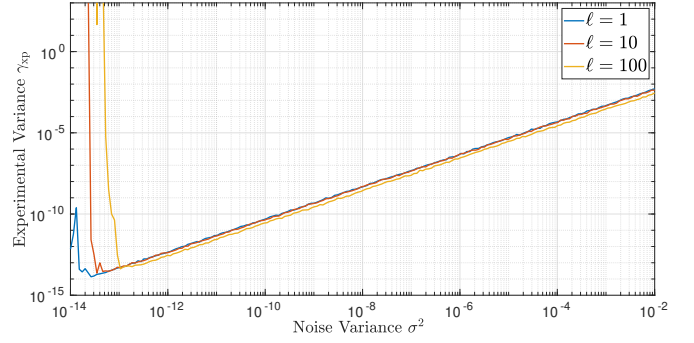- The size of the training dataset $K$.



Fig. 2. Evolution of the experimental forecasting variance as a function of the noise variance for three different values the forecasting sample index $\ell$.

In Sections IV-C1 and IV-C2, we study the influence of these parameters. A comparison with the theoretical results of Section III is also available.

*1) Sum of sine waves:* We proved that the linear dynamic model is sufficient to catch the dynamical behavior of signals taking the form (11). In order to validate this theoretical result, we apply the forecasting Algorithm 2 to a large number of realizations of the random vector $\mathbf{x}$ of size $N = 10^4$, following model (13), and such that the deterministic component $\mathbf{z}$ takes the form:

$$\mathbf{z}[n] = \cos\left(2\pi p_1 \frac{n}{M}\right) + A\cos\left(2\pi p_2 \frac{n}{M}\right), \quad \forall n \in \{1, \ldots, N\},$$

with $M = 150$, $p_1 = 10$, $p_2 = 33$ and $A = 1.4$. Besides, the additive noise is chosen to be Gaussian: $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.

*a) Influence of the Noise Variance $\sigma^2$:* Here, the size of the training dataset is set to $K = 450$. Then, the forecasting algorithm is run on 500 realizations on the discrete signal $\mathbf{x}$ for 200 different values of $\sigma$, logarithmically equispaced from $10^{-7}$ to $10^{-1}$. For each of these values, we determine the experimental bias, denoted as $\mu_{\text{xp}}[N-1+\ell]$, and experimental variance, denoted as $\gamma_{\text{xp}}[N-1+\ell, N-1+\ell]$. Fig. 2 shows the experimental forecasting variance as a function of the noise variance for three different values the forecasting sample index: $\ell = 1$, $\ell = 10$, and $\ell = 100$.

First, when the noise is weak, the variance of the forecasting estimator is dramatically high. For instance, in this specific example, the estimator $\tilde{\mathbf{x}}[N-1+\ell]$ for $\ell = 10$ is inaccurate as soon as the noise variance $\sigma^2$ is lower than $4 \times 10^{-14}$ (see the red line in Fig. 2). This validates the result of Theorem 1, highlighting the necessary presence of noise to ensure the numerical stability of the estimator. Furthermore, as soon as $\sigma^2$ is higher than this critical threshold, the estimator variance increases linearly with $\sigma^2$, as predicted by the bound (21) provided in Theorem 1. Note that the parameter $\ell$ has little influence on the quality of the estimator. This is due to the *stationarity* of the studied signal, allowing long-range forecasts without deterioration of the estimator.

*b) Influence of the Training Dataset Size $K$:* Here, the noise variance $\sigma$ is set to $\sigma = 10^{-2}$. Then, SigExt is run on 500 realizations of the discrete signal $\mathbf{x}$ for 200 dif-
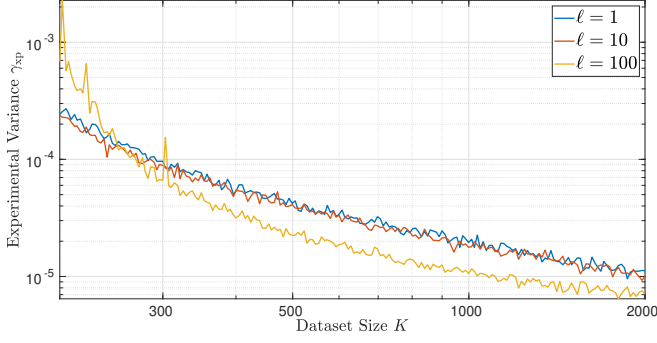
Fig. 3. Evolution of the experimental forecasting variance in function of the dataset size for three different values the forecasting sample index $\ell$.

| Extension method | CPU time (s) | MSE (mean $\pm$ SD) | Index $D$ of the STFT |
|---|---|---|---|
| SigExt ($M = 100$) | 0.008 | $1.133 \pm 0.077$ | $0.0091 \pm 0.0019$ |
| SigExt ($M = 750$) | 0.122 | $0.479 \pm 0.166$ | $0.0056 \pm 0.0019$ |
| SigExt ($M = 1500$) | 0.634 | $0.907 \pm 0.377$ | $0.0065 \pm 0.0023$ |
| Symmetric | $< 0.001$ | $5.452 \pm 0.002$ | $1.8209 \pm 0.0025$ |
| EDMD | 1.587 | $0.983 \pm 0.004$ | $0.1145 \pm 0.0018$ |
| GPR | 515.787 | $0.479 \pm 0.166$ | $0.0119 \pm 0.0037$ |
| TBATS | 1776.778 | $0.686 \pm 0.467$ | $0.3847 \pm 0.1397$ |

ferent values of training dataset size $K$, logarithmically equispaced from 450 to 2000. For each of these values, we determine the experimental bias $\mu_{\mathrm{xp}}[N-1+\ell]$ and variance $\gamma_{\mathrm{xp}}[N-1+\ell, N-1+\ell]$. Fig. 3 shows the experimental forecasting variance as a function of the training dataset size for three different values the forecasting sample index: $\ell = 1$, $\ell = 10$, and $\ell = 100$.

As anticipated by Theorem 1, the variance of the estimator decreases linearly with $\frac{1}{K}$ as long as the dataset size is large enough. Indeed, in the log-log plot displayed in Fig. 3, as soon as $K > 700$, each of the three curves representing the forecasting variance as a function of $K$ tends towards an asymptote of slope $-1$. In practice, the asymptotic behavior of SigExt, illustrated here on a synthetic signal, can be generalized to the study of real-life signals, as long as the duration of the recorded signal largely exceeds that of the desired extension, which allows the user to set $K \gg L$.

This study overlooks the analysis of the influence of the parameter $M$, whose influence on the value of the experimental variance is numerically not significant as long as $M \ll 2K$ and $\ell < M$. The choice of this parameter is especially crucial when the deterministic component of the signal is no longer stationary. The AHM, discussed below, is an example.

*2) Adaptive Harmonic Model:* Consider a signal satisfying the AHM so that the instantaneous frequencies and amplitudes of its components vary over time. The deterministic component $\mathbf{z}$ of the random vector $\mathbf{x}$, constructed following the model (13), takes the following form, for all $n \in \{1, \dots, N\}$:

$$\mathbf{x}[n] = \cos\left(2\pi\phi_1[n]\right) + R[n]\cos\left(2\pi\phi_2[n]\right) ,$$

where the instantaneous amplitude $R$ is given by:

$$R[n] = 1.4 + 0.2\cos\left(4\pi\frac{n}{N}\right) ,$$

and the instantaneous phases are such that:

$$\phi_1[n] = \frac{p_1}{P}\left(n + \frac{0.01}{2\pi}\cos\left(2\pi\frac{n}{N}\right)\right)$$
$$\phi_2[n] = p_2\frac{n}{P} + \frac{20}{2Nf_{\mathsf{s}}}n^2 .$$

Besides, the noise is chosen to be Gaussian, $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, and we take $N = 10^4$, $P = 750$, $p_1 = 10$, $p_2 = 23$.

*a) Influence of the subsignals length $M$:* To evaluate the forecasting quality, SigExt is applied to 1000 realizations of the above-described signal. We forecast signal extensions of 0.1 s, i.e., $L = 700$. Table I shows the resulting mean and standard deviation (SD) of the experimental MSE, according to three different values of the subsignals length $M$. Note that choosing too small an $M$ does not provide enough information to forecast the signal satisfactorily. Furthermore, too large an $M$ makes the algorithm insensitive to local nonstationarities produced by frequency variations. The user must then make a compromise on the choice of $M$ in order to optimize the performance of the algorithm. A choice of $M$ such that the subsignals contain about a dozen oscillations is generally appropriate.

*b) Comparison of the extensions methods:* The same set of experiments is run on the extension methods outlined in Section IV-A. The results, displayed in Table I, show that the naive extension we propose gives satisfying results. Indeed, when the subsignal length $M$ is optimally chosen, SigExt performance is of the same order as that of ore sophisticated methods, like GPR or TBATS. Besides, even though these methods could be slightly more robust to perturbations, they are substantially limited by the computing time they require, which prevents them from being used to exploit real-time data. Thus, SigExt is the extension method that optimizes the trade-off between forecasting quality and computing time. Besides, the last column of Table I gives the performance index $D$ of the boundary-free STFT associated with each extension method. This illustrates the ability of our algorithm to reduce boundary effects to a level comparable with what is possible with the GPR or TBATS extensions. The performance of BoundEffRed on other TF representations is more thoroughly analyzed on real-life signals in the following.

### D. Real Physiological Signals

*1) Respiratory Signal:* We first consider a thoracic movement signal recorded by the piezo-electrical sensor of length 6 hours and 20 minutes. The signal is sampled at $f_{\mathsf{s}} = 100$ Hz. A small portion of this signal is displayed in Fig. 4.
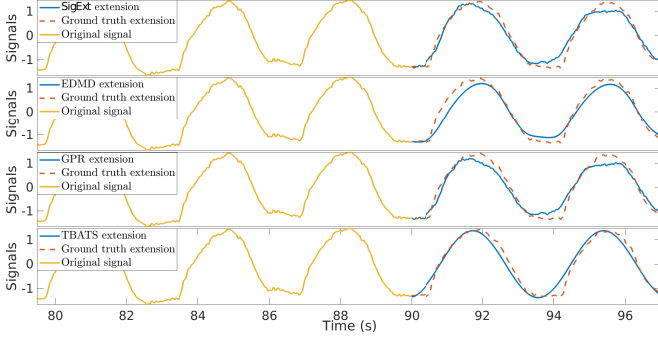
Fig. 4. Extended respiratory signal (blue) obtained by the SigExt forecasting (first panel), the EDMD forecasting (second panel), the GPR forecasting (third panel), and the TBATS forecasting (fourth panel), superimposed with the ground truth signal (red-dashed line).

TABLE II
RESPIRATORY SIGNAL: PERFORMANCE OF THE EXTENSION METHODS AND THE ASSOCIATED BOUNDARY-FREE TF REPRESENTATIONS

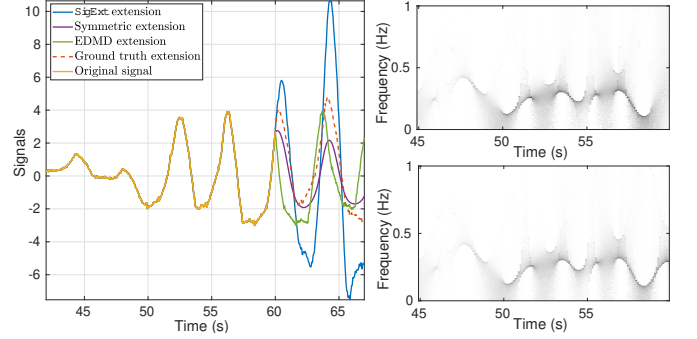| Extension method | MSE | Performance index $D$ (mean $\pm$ SD) | | | |
|---|---|---|---|---|---|
| | | STFT | SST | RS | ConceFT |
| SigExt | 0.292 $\pm$4.438 | 0.370 $\pm$0.623 | 0.408 $\pm$0.436 | 0.866 $\pm$0.879 | 0.423 $\pm$0.344 |
| Symmetric | 0.044 $\pm$0.111 | 1.162 $\pm$0.893 | 1.173 $\pm$0.886 | 1.022 $\pm$0.281 | 1.144 $\pm$0.579 |
| EDMD | 0.026 $\pm$0.112 | 0.359 $\pm$0.266 | 0.422 $\pm$0.282 | 0.828 $\pm$0.248 | 0.449 $\pm$0.296 |
| GPR | 0.331 $\pm$4.858 | 0.391 $\pm$0.853 | 0.411 $\pm$0.406 | 0.897 $\pm$1.140 | 0.430 $\pm$0.364 |



Fig. 5. Extensions of a segment of the respiratory signal (left) where SigExt is outperformed by the EDMD and Symmetric extensions. Corresponding SST (top-right) and boundary-free SST obtained with SigExt (bottom-right).

From that long signal, we build a dataset of 379 nonoverlapping signals of 60 seconds, i.e., $N = 6000$. For each segment, we apply the forecasting methods introduced in Section IV-C2, including the SigExt method detailed in Algorithm 2. We forecast 7-second-long extensions on each segment of the signal, corresponding to $L = 700$. Thus, in order to catch the dynamical behavior, the size of the training signal $M$ is chosen so that $M = \lfloor 1.5L \rfloor$. As a result of Section IV-C1, we take: $K = \lfloor 2.5M \rfloor$. The extensions obtained on one of these subsignals are shown in Fig. 4. The resulting MSEs of different methods are given in Table II. Note that the results are given in "mean $\pm$ standard deviation" format. The TBATS extension method is not implemented, as its excessive computing time makes it impossible to implement it on 6000 segments within a reasonable period of time. Note that the MSE of SigExt is on average higher than the MSE of EDMD or the symmetric extension, with a huge standard deviation. This variability is caused by the presence of few segments contaminated by artifacts so that it is unpredictable via a too simple dynamical model like (4). The left of Fig. 5 illustrates one of these outliers, where SigExt fails to catch the fast varying dynamic of the instantaneous amplitude to satisfactorily forecast the signal. The EDMD and symmetric extensions are more robust to those situations, as shown in this example, and in Table II. Nevertheless, SigExt provides a sufficiently relevant extension to give TF representations sparingly affected by boundary effects. On the right of Fig. 5, we display a comparison between the right boundary of SST of the same segment of signal (top-right), and its boundary-free SST obtained after the SigExt forecasting (bottom-right). The extension of the instantaneous frequency visible on the right side of the image illustrates the reduction of boundary effects produced despite an inaccurate signal forecasting.

We then apply BoundEffRed for diverse TF representations: STFT, SST, RS, as well as *Concentration of Frequency and Time* (ConceFT), a generalized multitaper SST-based representation introduced in [33]. We mention that unlike other representations, the standard RS is

not causal and rigorously requires knowing the STFT on the whole time-frequency plane before making any reassignment. To enable real-time implementation of RS, we use here a real-time version of RS proposed in [35].

We set the extension length $L$ accordingly to the window length used by the TF analysis tool. For instance, the window length used to evaluate the STFT is of 1400 samples. To prevent the STFT from being sensitive to the boundaries, we set $L = 700$. In this way, the evaluation of the spectral content of the signal near its boundary is not limited by a lack of information in the interval delimited by the window support. From now on, all results are given for $L$ equal to the half of the width of the window used in the TF transform. In Table II, we give the averaged performance index (25) by evaluating the whole TF representations (including boundaries). Even though SigExt performs moderately well in the sense of MSE, the boundary effects are dramatically reduced on the TF representations, and the averaged performance is in the same order of that given by EDMD or GPR. A t-test is performed to compare the performance index of SigExt with those of other methods. Under the null hypothesis that the means are equal, with a 5% significance level, the t-tests show no statistically significant difference between SigExt and EDMD or GPR, regardless of the representation considered.

*2) Two-Component Respiratory Signal:* Then, we consider a second respiratory signal that contains not only the respiratory cycle, but also a cardiac component
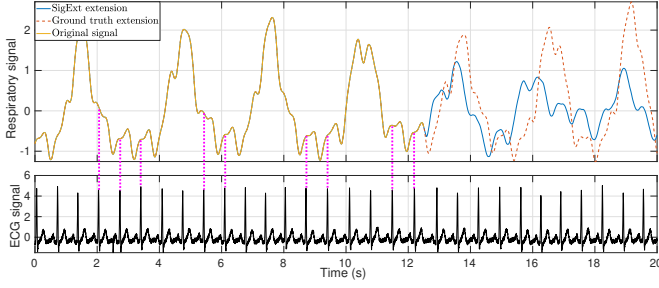
Fig. 6. Two-component respiratory signal (top, yellow), and the associated SigExt extension (blue); the simultaneously recorded ECG signal is below. The vertical magenta dotted lines indicate the cardiac cycles in the respiratory signal.

known as the cardiogenic artifact [36]. The top of Fig. 6 shows an excerpt of the measured thoracic impedance-based respiratory signal, recorded during a broncho-scopic sedation procedure, along with the SigExt extension. Below, the simultaneously recorded ECG signal is depicted. The magenta dotted lines emphasize the coincidences between the cardiogenic artifact of the respiratory signal and the QRS complexes of the ECG signal. Such cardiac information can be harvested and utilized when other first-line heart rate information resource is not available or broken [37].

The top of Fig. 7 displays the ordinary SST and the boundary-free SST of this respiratory signal, sampled at 64 Hz. This TF analysis brings out both components—the respiratory component, indicated by the green arrows, is located around the fundamental frequency 0.3 Hz and its multiples, while the cardiac component, indicated by the blue arrows, is located around 1.5 Hz. Note that, here, boundary effects are only reduced on the right-side boundary of the TF domain, delimited by the red dashed boxes. The middle of Fig. 7 show a zoom near the right-boundary of the ordinary SST, while the bottom of Fig. 7 respectively show the respective zoom on the boundary-free SST. In this area, the boundary-free SST makes it possible to disentangle the components contained in the signal. Moreover, the performance index of this representation takes the value $D = 0.705$. This means that BoundEffRed has reduced the right-side boundary effects by about 30% with respect to the ordinary SST. This shows the ability of our algorithm to work on signals containing several nonstationary components.

*3) Photoplethysmogram:* We consider a 640-second photoplethysmogram (PPG) signal extracted from the Physionet dataset [38], [39], sampled at $f_s = 125$ Hz. A portion of this signal is displayed in Fig. 8. The estimated 5-second extensions of this segment obtained by SigExt, EDMD, and GPR forecastings are superimposed to the ground-truth extension in Fig. 8.

We divide the signal into 32-second-long segments and apply BoundEffRed to each of them. Table III shows the performance index $D$ of the boundary-free TF representations, averaged over the signals. For all the TF representations considered, the results show that SigExt
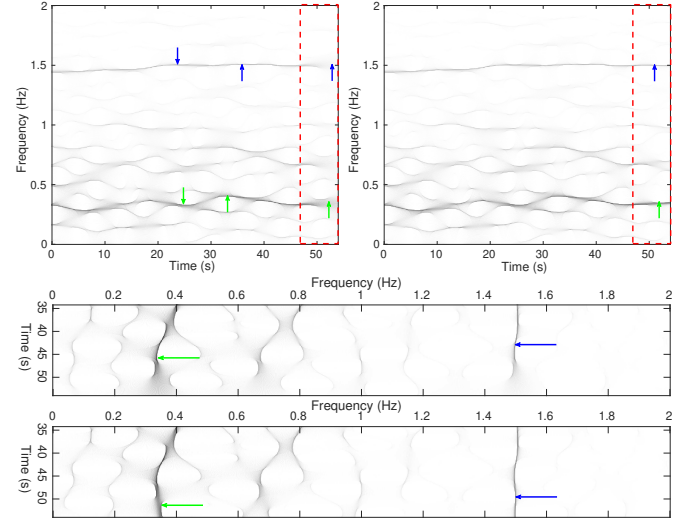


Fig. 7. Ordinary SST (top-left) and boundary-free SST (top-right) of a two-component respiratory signal. Middle and bottom: zoom on the areas delimited by the red dashed rectangles; these images are rotated 90 degrees clockwise. The window length for the SSTs is 20 seconds. The instantaneous frequency of the respiration is indicated by the green arrows, while the instantaneous frequency of the cardiogenic artifact is indicated by the blue arrows.
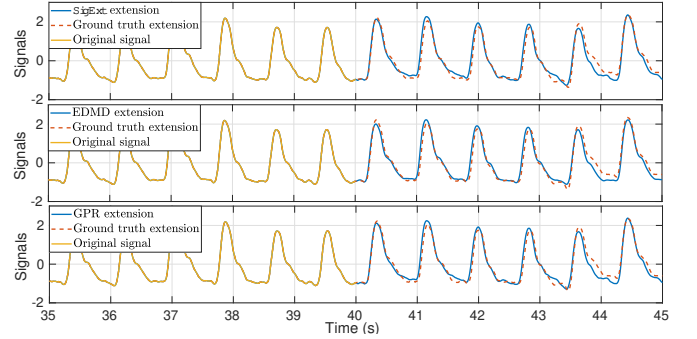


Fig. 8. Extended PPG signal (blue) obtained by the SigExt forecasting (top), the EDMD forecasting (middle), and the GPR forecasting (bottom), superimposed with the ground truth signal (red dash).

reduces boundary effects about as efficiently as the extensions given by EDMD or GPR. As for the respiratory signal, t-tests show no statistically significant difference between SigExt and EDMD or GPR, regardless of the representation considered. For visual inspection, the TF representation of SST resulting from the BoundEffRed strategy is shown in the bottom-right panel of Fig. 1, where SST is applied to the portion of PPG displayed in Fig. 8. It produces a significant improvement in the quality of the SST near boundaries. Indeed, the blurring visible when zooming on the right boundary of the SST has almost vanished. Real-time tracking of the instantaneous frequencies contained in the measured signal is therefore greatly facilitated.

To further evaluate the influence of the noise level on the performance of BoundEffRed, we artificially add Gaussian noise to the measured PPG signal. It is thus an additional noise to the measurement noise actually

TABLE III
PPG Signal: Performance of the Boundary-Free TF
Representations According to the Extension Method

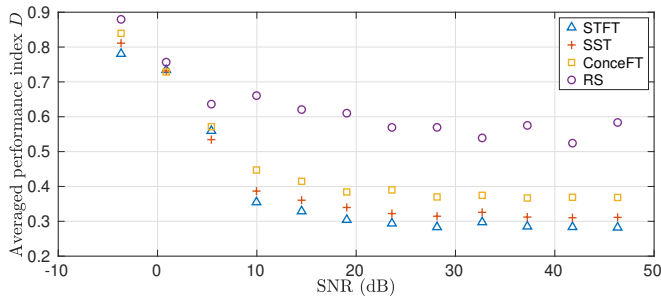| Extension method | MSE | Performance index $D$ (mean $\pm$ SD) | | | |
|---|---|---|---|---|---|
| | | STFT | SST | ConceFT | RS |
| SigExt | 0.018 $\pm$0.014 | 0.280 $\pm$0.107 | 0.309 $\pm$0.112 | 0.367 $\pm$0.183 | 0.534 $\pm$0.160 |
| Symmetric | 0.037 $\pm$0.007 | 1.168 $\pm$0.390 | 1.209 $\pm$0.340 | 1.310 $\pm$0.140 | 0.983 $\pm$0.304 |
| EDMD | 0.012 $\pm$0.005 | 0.289 $\pm$0.126 | 0.319 $\pm$0.134 | 0.375 $\pm$0.163 | 0.503 $\pm$0.163 |
| GPR | 0.018 $\pm$0.013 | 0.276 $\pm$0.106 | 0.303 $\pm$0.110 | 0.361 $\pm$0.165 | 0.544 $\pm$0.157 |



Fig. 9. Performance index of BoundEffRed applied on a PPG, in function of the SNR.

contained in the signal. Fig. 9 shows the average performance index of BoundEffRed for different values of the Signal-to-Noise Ratio (SNR). Whatever the representation considered, BoundEffRed is relatively robust to noise, as long as the SNR remains above 10 dB. Below this level, the reduction of boundary effects gradually deteriorates.

*E. Real-Time Implementation*

We simulate the real-time implementation of the SST from BoundEffRed applied to the PPG described in Section IV-D3, subsampled by a factor of 2. The test is performed on a 2-Core Intel Core i5 CPU running at 1.7 GHz and 7.7 GB of RAM.

For this signal, sampled at $f_{\mathsf{s}} = 65.5$ Hz, the suitable window is 8 seconds long. Therefore, the extension must be over 4 seconds, i.e., $L = 250$. The forecasting step in each iteration of BoundEffRed then takes no more than $t_{\text{forecast}} = 46$ ms. Denote by $H \geq 1$ the hop size in samples, that is the number of samples between successive columns of the SST. Besides, since the support of the analysis window covers $2L$ samples, the update of the SST requires the calculation of $\lceil 2L/H \rceil$ new columns of the SST. In this example, the frequency dimension of the SST being 512, the computational time of one column is $t_{\text{SST}} = 2.08$ ms, on average. A general rule to determine the acceptable values of $H$ for real-time implementation is verifying that the computational time to update the boundary-free TF representation is smaller than the lag between $H$ samples; that is,

$$t_{\text{forecast}} + \left\lceil \frac{L}{H} \right\rceil t_{\text{SST}} < \frac{H}{f_{\mathsf{s}}} \ .$$

In our example, taking $H \geq 8$ samples is sufficient to ensure the feasibility of real-time implementation. This value is reasonable because it allows a maximum overlap of 98.4% of the window length. SST variations are therefore perceived gradually. Combined with the low latency of the algorithm, this allows for smooth monitoring of the real-time SST. A video clip illustrating this is available on GitHub, at https://github.com/AdMeynard/BoundaryEffectsReduction/tree/master/Animations.

## V. Conclusion

In this paper, we propose an algorithm, named Bound-EffRed, for the real-time reduction of boundary effects in TF representations. This method is based on an extension of the signal obtained by a simple-minded and numerically efficient forecasting. We have shown theoretically that the chosen dynamic model is sufficient to extend signals formed by a sum of sine waves. Moreover, the low computational time allows us to switch to a real-time implementation of BoundEffRed, unlike other existing forecasting methods. The numerical results also confirmed the robustness to noise of BoundEffRed, as well as its ability to be applied to many TF representations. Additional applications to ECG and PPG are included in the Supplementary Materials.

Various improvements can be considered to make the algorithm more robust. In particular, we have noticed (see Fig. 5) that when the regular oscillations of the observed signal break, the forecasting step is no longer relevant, and only slows down the calculation of the TF representation. A preliminary step should then be added to the algorithm to detect signal activity and disable the forecasting step when possible. More fundamentally, one can also consider accelerating the computational time by optimizing the forecasting step to improve the real-time performance of BoundEffRed. Indeed, each new forecast requires the inversion of the matrix $\mathbf{XX}^T$ of size $M \times M$. However, the matrix $\mathbf{X}$ used to forecast $\mathbf{x}_{N+H}$ at the current iteration differs from the one used at the previous iteration to forecast $\mathbf{x}_N$ by only $H$ columns. It then seems natural to take inspiration from the work of Strobach [40], generalized by Badeau *et al.* [41], which proposes a fast algorithm for the singular value decomposition of successive data matrices taking the same form as $\mathbf{X}$. Their algorithm is limited to the case where $H = 1$. Developing an extension of this algorithm to variations of $H > 1$ columns would then allow us to efficiently update the matrix $\left(\mathbf{XX}^T\right)^{-1}$. These avenues will be explored in our future work.

## References

[1] D. Stowell, *Computational Analysis of Sound Scenes and Events*. Springer, 2018, ch. Computational Bioacoustic Scene Analysis, pp. 303–333.

[2] M. Müller, D. P. W. Ellis, A. Klapuri, and G. Richard, "Signal processing for music analysis," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 6, pp. 1088–1110, 2011.

[3] Z. Peng, F. Chu, and Y. He, "Vibration signal analysis and feature extraction based on reassigned wavelet scalogram," *Journal of Sound and Vibration*, vol. 253, no. 5, pp. 1087 – 1100, 2002. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0022460X01940854

[4] M. Akay, *Detection and Estimation Methods for Biomedical Signals*, 1st ed. USA: Academic Press, Inc., 1996.

[5] P. Stoica and R. Moses, *Spectral analysis of signals*. Upper Saddle River, N.J.: Pearson/Prentice Hall, 2005.

[6] G. Matz, F. Hlawatsch, and W. Kozek, "Generalized evolutionary spectral analysis and the Weyl spectrum of nonstationary random processes," *IEEE Transactions on Signal Processing*, vol. 45, no. 6, pp. 1520–1534, Jun. 1997.

[7] K. Gröchenig, *Foundations of time-frequency analysis*, ser. Applied and Numerical Harmonic Analysis. Boston, MA: Birkhäuser Inc., 2001.

[8] P. Flandrin, *Time-frequency/time-scale analysis*, ser. Wavelet Analysis and its Applications. San Diego: Academic Press Inc., 1999, vol. 10.

[9] I. Daubechies, *Ten lectures on wavelets*, ser. CBMS-NSF Regional Conference Series in Applied Mathematics. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM), 1992, vol. 61.

[10] I. Daubechies, J. Lu, and H.-T. Wu, "Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool," *Applied and Computational Harmonic Analysis*, vol. 30, no. 2, pp. 243 – 261, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1063520310001016

[11] F. Auger, P. Flandrin, Y.-T. Lin, S. McLaughlin, S. Meignen, T. Oberlin, and H.-T. Wu, "Time-frequency reassignment and synchrosqueezing: An overview," *IEEE Signal Processing Magazine*, vol. 30, no. 6, pp. 32–41, 2013.

[12] N. Delprat, B. Escudié, P. Guillemain, R. Kronland-Martinet, P. Tchamitchian, and B. Torrésani, "Asymptotic wavelet and Gabor analysis: extraction of instantaneous frequencies," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 644–664, Mar. 1992.

[13] C. K. Chui and E. Quak, *Numerical Methods in Approximation Theory*, ser. ISNM 105: International Series of Numerical Mathematics. Basel: Birkäuser, 1992, vol. 9, ch. Wavelets on a Bounded Interval, pp. 53–75.

[14] U. Depczynski, K. Jetter, K. Molt, and A. Niemöller, "The fast wavelet transform on compact intervals as a tool in chemometrics: II. Boundary effects, denoising and compression," *Chemometrics and Intelligent Laboratory Systems*, vol. 49, no. 2, pp. 151 – 161, 1999. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0169743999000374

[15] I. Kharitonenko, X. Zhang, and S. Twelves, "A wavelet transform with point-symmetric extension at tile boundaries," *IEEE Transactions on Image Processing*, vol. 11, no. 12, pp. 1357–1364, 2002.

[16] L. Chen, T. Q. Nguyen, and K.-P. Chan, "Symmetric extension methods for m-channel linear-phase perfect-reconstruction filter banks," *IEEE Transactions on Signal Processing*, vol. 43, no. 11, pp. 2505–2511, 1995.

[17] J. R. Williams and K. Amaratunga, "A discrete wavelet transform without edge effects using wavelet extrapolation," *The Journal of Fourier Analysis and Applications*, vol. 3, pp. 435–449, 1997.

[18] M. O. Williams, I. G. Kevrekidis, and C. W. Rowley, "A Data-Driven Approximation of the Koopman Operator: Extending Dynamic Mode Decomposition," *Journal of Nonlinear Science*, vol. 25, pp. 1307–1346, 2015.

[19] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, ser. Adaptive Computation and Machine Learning series. The MIT Press, 2006.

[20] S. Roberts, M. Osborne, M. Ebden, S. Reece, N. Gibson, and S. Aigrain, "Gaussian processes for time-series modelling," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 371, no. 1984, p. 20110550, 2013. [Online]. Available: https://royalsocietypublishing.org/doi/abs/10.1098/rsta.2011.0550

[21] J. Vargas and S. McLaughlin, "Speech analysis and synthesis based on dynamic modes," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 8, pp. 2566–2578, 2011.

[22] C. Grebogi, S. M. Hammel, J. A. Yorke, and T. Sauer, "Shadowing of physical trajectories in chaotic dynamics: Containment and refinement," *Phys. Rev. Lett.*, vol. 65, pp. 1527–1530,

Sep 1990. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.65.1527

[23] A. M. D. Livera, R. J. Hyndman, and R. D. Snyder, "Forecasting time series with complex seasonal patterns using exponential smoothing," *Journal of the American Statistical Association*, vol. 106, no. 496, pp. 1513–1527, 2011. [Online]. Available: https://doi.org/10.1198/jasa.2011.tm09771

[24] M. West and J. Harrison, *Bayesian forecasting and dynamic models*. Springer Science & Business Media, 2006.

[25] J. Fan and I. Gijbels, *Local polynomial modelling and its applications: monographs on statistics and applied probability 66*. CRC Press, 1996, vol. 66.

[26] P. Hall and J. D. Opsomer, "Theory for penalised spline regression," *Biometrika*, vol. 92, no. 1, pp. 105–118, 2005.

[27] Y.-W. Chang, C.-J. Hsieh, K.-W. Chang, M. Ringgaard, and C.-J. Lin, "Training and testing low-degree polynomial data mappings via linear svm." *Journal of Machine Learning Research*, vol. 11, no. 4, 2010.

[28] J. S. Marron, S. Adak, I. Johnstone, M. Neumann, and P. Patil, "Exact risk analysis of wavelet regression," *Journal of Computational and Graphical Statistics*, vol. 7, no. 3, pp. 278–309, 1998.

[29] P. R. Vlachas, W. Byeon, Z. Y. Wan, T. P. Sapsis, and P. Koumoutsakos, "Data-driven forecasting of high-dimensional chaotic systems with long short-term memory networks," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 474, no. 2213, p. 20170844, 2018.

[30] R. J. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice*. OTexts, 2018.

[31] P. J. Schmid, "Dynamic mode decomposition of numerical and experimental data," *Journal of Fluid Mechanics*, vol. 656, pp. 5–28, 2010.

[32] Y.-C. Chen, M.-Y. Cheng, and H.-T. Wu, "Non-parametric and adaptive modelling of dynamic periodicity and trend with heteroscedastic and dependent errors," *Journal of the Royal Statistical Society Series B*, vol. 76, no. 3, pp. 651–682, June 2014. [Online]. Available: https://ideas.repec.org/a/bla/jorssb/v76y2014i3p651-682.html

[33] I. Daubechies, Y. G. Wang, and H.-T. Wu, "ConceFT: concentration of frequency and time via a multitapered synchrosqueezed transform," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2065, p. 20150193, 2016. [Online]. Available: https://royalsocietypublishing.org/doi/abs/10.1098/rsta.2015.0193

[34] M. Korda and I. Mezić, "Linear predictors for nonlinear dynamical systems: Koopman operator meets model predictive control," *Automatica*, vol. 93, pp. 149 – 160, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S000510981830133X

[35] Y.-T. Lin and H.-T. Wu, "ConceFT for Time-Varying Heart Rate Variability Analysis as a Measure of Noxious Stimulation During General Anesthesia," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 1, pp. 145–154, 2017.

[36] T. C. Smith, A. Green, and P. Hutton, "Recognition of cardiogenic artifact in pediatric capnograms," *Journal of Clinical Monitoring*, vol. 10, no. 4, pp. 270–275, Jul. 1994.

[37] Y. Lu, H.-T. Wu, and J. Malik, "Recycling cardiogenic artifacts in impedance pneumography," *Biomedical Signal Processing and Control*, vol. 51, pp. 162 – 170, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1746809419300710

[38] M. A. F. Pimentel, A. E. W. Johnson, P. H. Charlton, D. Birrenkott, P. J. Watkinson, L. Tarassenko, and D. A. Clifton, "Toward a robust estimation of respiratory rate from pulse oximeters," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 8, pp. 1914–1923, 2017.

[39] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, M. R. G., M. J. E., M. G. B., P. C. K., and S. H. E., "Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. circulation [online]. 101 (23),," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[40] P. Strobach, "Square hankel svd subspace tracking algorithms," *Signal Processing*, vol. 57, no. 1, pp. 1 – 18, 1997. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S016516849600182X

[41] R. Badeau, G. Richard, and B. David, "Sliding window adaptive svd algorithms," *IEEE Transactions on Signal Processing*, vol. 52, no. 1, pp. 1–10, Jan 2004.

# Supplementary Materials for "An Efficient Forecasting Approach to Reduce Boundary Effects in Real-Time Time-Frequency Analysis"

Adrien Meynard, Hau-Tieng Wu

## I. Proof of Theorem 1

### A. Preliminaries

*1) Notations:* Consider $\mathbf{z} \in \mathbb{R}^N$ the deterministic signal defined in (11), and denote by $\mathbf{z}_k \in \mathbb{R}^M$ the subsignal such that

$$\mathbf{z}_k[m] = \mathbf{z}[N - K - M + k + m], \quad \forall m \in \{0, \dots, M-1\}, \; k \in \{0, \dots, K\} \,.$$

Define $\mathbf{Z} \in \mathbb{R}^{M \times K}$ and $\mathbf{Z}' \in \mathbb{R}^{M \times K}$, the matrices such that

$$\mathbf{Z} \triangleq \begin{pmatrix} \mathbf{z}_0 & \cdots & \mathbf{z}_{K-1} \end{pmatrix} \,,$$

$$\mathbf{Z}' \triangleq \begin{pmatrix} \mathbf{z}_1 & \cdots & \mathbf{z}_K \end{pmatrix} \,.$$

Let $\mathbf{D} \in \mathbb{R}^{M \times M}$ be the matrix defined by $\mathbf{D}[m, m'] = \delta_{m+1, m'}$. Recall the model (13). Based on the definition of matrices $\mathbf{X}$ and $\mathbf{Y}$, we have:

$$\frac{1}{K}\mathbf{X}\mathbf{X}^T = \underbrace{\frac{1}{K}\mathbf{Z}\mathbf{Z}^T + \sigma^2\mathbf{I}}_{\triangleq \mathbf{S}^{(0)}} + \mathbf{E}^{(0)} \tag{26}$$

$$\frac{1}{K}\mathbf{Y}\mathbf{X}^T = \underbrace{\frac{1}{K}\mathbf{Z}'\mathbf{Z}^T + \sigma^2\mathbf{D}}_{\triangleq \mathbf{S}^{(1)}} + \mathbf{E}^{(1)} \,, \tag{27}$$

where $\mathbf{E}^{(a)} \triangleq \sigma\mathbf{E}_1^{(a)} + \sigma^2\mathbf{E}_2^{(a)}$,

$$\mathbf{E}_1^{(a)}[m, m'] = \frac{1}{K}\sum_{k=0}^{K-1} \mathbf{z}[N_0 + m + a + k]\mathbf{w}[N_0 + m' + k] + \mathbf{w}[N_0 + m + a + k]\mathbf{z}[N_0 + m' + k] \,,$$

and

$$\mathbf{E}_2^{(a)}[m, m'] = \frac{1}{K}\sum_{k=0}^{K-1} \mathbf{w}[N_0 + m + a + k]\mathbf{w}[N_0 + m' + k] - \delta_{(m+a)m'} \,,$$

with $a \in \{0, 1\}$. We call $\mathbf{E}^{(0)}$ and $\mathbf{E}^{(1)}$ *error matrices* because:

$$\mathbb{E}\{\mathbf{E}^{(0)}\} = \mathbb{E}\{\mathbf{E}_1^{(0)}\} = \mathbb{E}\{\mathbf{E}_2^{(0)}\} = \mathbf{0}$$

$$\mathbb{E}\{\mathbf{E}^{(1)}\} = \mathbb{E}\{\mathbf{E}_1^{(1)}\} = \mathbb{E}\{\mathbf{E}_2^{(1)}\} = \mathbf{0} \,.$$

Thus, the random matrix $\tilde{\mathbf{A}}$, defined in equation (8), is expressed in function of the above-defined matrices as:

$$\tilde{\mathbf{A}} = \left(\mathbf{S}^{(1)} + \mathbf{E}^{(1)}\right)\left(\mathbf{S}^{(0)} + \mathbf{E}^{(0)}\right)^{-1} \,.$$

Define $\mathbf{A}_0$ the deterministic matrix such that

$$\mathbf{A}_0 \triangleq \mathbf{S}^{(1)}\mathbf{S}^{(0)^{-1}} \,. \tag{28}$$

A. Meynard is with the Department of Mathematics, Duke University, Durham, NC, 27708 USA.

H.-T. Wu is with the Department of Mathematics and Department of Statistical Science, Duke University, Durham, NC, 27708 USA; Mathematics Division, National Center for Theoretical Sciences, Taipei, Taiwan.

A. Meynard is the corresponding author (e-mail: adrien.meynard@duke.edu).

We denote by $\boldsymbol{\alpha}_0^{(\ell)}$ the last row of $\mathbf{A}_0^\ell$. As a result, for $\ell \in \mathbb{N}^*$, the error vector $\mathbf{h}^{(\ell)}$ defined by $\mathbf{h}^{(\ell)} \overset{\Delta}{=} \boldsymbol{\alpha}^{(\ell)} - \boldsymbol{\alpha}_0^{(\ell)}$ satisfy the equation

$$
\begin{aligned}
\mathbf{h}^{(\ell)} &= \mathbf{e}_M^T \left( \tilde{\mathbf{A}}^\ell - \mathbf{A}_0^\ell \right) \\
&= \mathbf{e}_M^T \left( \left( (\mathbf{S}^{(1)} + \mathbf{E}^{(1)})(\mathbf{S}^{(0)} + \mathbf{E}^{(0)})^{-1} \right)^\ell - \mathbf{A}_0^\ell \right) .
\end{aligned}
\tag{29}
$$

The randomness of $\mathbf{h}^{(\ell)}$ completely comes from the error matrices. Besides, notice that the first $M-1$ rows in $\mathbf{E}^{(1)}$ equal to the last $M-1$ rows of $\mathbf{E}^{(0)}$. We gather all sources of randomness into a vector $\mathbf{g} \in \mathbb{R}^{M(M+1)}$, defined as

$$
\mathbf{g} = \mathrm{vec}\left( \begin{bmatrix} \mathbf{E}^{(0)} \\ \mathbf{e}_M^T \mathbf{E}^{(1)} \end{bmatrix} \right) ,
\tag{30}
$$

where "vec" denotes the vectorization operator, that concatenates the columns of a given matrix on top of one another. Then, we can write $\mathbf{h}^{(\ell)}$ as $\mathbf{h}^{(\ell)} = f^{(\ell)}(\mathbf{g})$ where $f^{(\ell)}$ is a deterministic function such that:

$$
\begin{aligned}
f^{(\ell)} &: \mathbb{R}^{M(M+1)} \to \mathbb{R}^M \\
\mathbf{g} &\mapsto \mathbf{h}^{(\ell)} .
\end{aligned}
\tag{31}
$$

In the following paragraph, we provide some useful lemmas to prove Theorem 1.

*2) Lemmas:*

**Lemma 1** (Expressions of $\mathbf{A}_0$ and $\mathbf{S}^{(0)^{-1}}$). *Let $\mathbf{S}^{(0)}$ be the $M \times M$ matrix defined in (26). Let $\mathbf{A}_0$ the $M \times M$ matrix defined in (28). Assume the deterministic signal $\mathbf{z}$ takes the form (11), and the observed noisy signal takes the form (13). Then, the inverse of the matrix $\mathbf{S}^{(0)}$ is given by*

$$
\mathbf{S}^{(0)^{-1}}[m, m'] = \frac{1}{\sigma^2}\delta_{m,m'} - \frac{2}{M\sigma^2}\sum_{j=1}^{J}\frac{1}{1 + \frac{4\sigma^2}{M\Omega_j^2}}\cos\left(2\pi p_j\frac{m-m'}{M}\right) ,
\tag{32}
$$

*and the matrix $\mathbf{A}_0$ is given by*

$$
\mathbf{A}_0[m, m'] = \delta_{m+1,m'} + \frac{2}{M}\sum_{j=1}^{J}\frac{1}{1 + \frac{4\sigma^2}{M\Omega_j^2}}\cos\left(2\pi p_j\frac{m'}{M}\right)\delta_{m+1,M} .
\tag{33}
$$

*Let $\|\cdot\|_{\max}$ denote the maximum norm of a matrix, i.e., $\|\mathbf{M}\|_{\max} = \max_{n,n'}|\mathbf{M}[n,n']|$. Then,*

$$
\left\| \mathbf{S}^{(0)^{-1}} \right\|_{\max} \leq \frac{1}{\sigma^2}\left(1 + \frac{2J}{M}\right) ,
\tag{34}
$$

$$
\|\mathbf{A}_0\|_{\max} \leq \max\left(1, \frac{2J}{M}\right) .
\tag{35}
$$

*Proof.* It follows from the signal model (11) that the matrices $\mathbf{S}^{(0)}$ and $\mathbf{S}^{(1)}$ take the following form:

$$
\begin{aligned}
\mathbf{S}^{(a)}[m, m'] &= \sigma^2\delta_{(m+a),m'} + \sum_{j,j'=1}^{J}\frac{\Omega_j\Omega_{j'}}{K}\sum_{k=0}^{K-1}\cos\left(2\pi\frac{f_j}{f_s}(N_0 + m + a + k) + \varphi_j\right)\cos\left(2\pi\frac{f_{j'}}{f_s}(N_0 + m' + k) + \varphi_{j'}\right) \\
&= \sigma^2\delta_{(m+a),m'} + \sum_{j=1}^{J}\frac{\Omega_j^2}{2K}\sum_{k=0}^{K-1}\cos\left(2\pi\frac{f_j}{f_s}(m+a-m')\right) + \cos\left(2\pi\frac{f_j}{f_s}(2k + m + a + m' + 2N_0)\right) \\
&= \sigma^2\delta_{(m+a),m'} + \sum_{j=1}^{J}\left(\frac{\Omega_j^2}{2}\cos\left(2\pi\frac{f_j}{f_s}(m+a-m')\right) + \frac{\Omega_j^2}{2K}\underbrace{\sum_{k=0}^{K-1}\cos\left(2\pi\frac{f_j}{f_s}(2k + m + a + m' + 2N_0)\right)}_{=0 \text{ because } \frac{f_j}{f_s} = \frac{p'_j}{K}}\right) \\
&= \sigma^2\delta_{(m+a),m'} + \sum_{j=1}^{J}\frac{\Omega_j^2}{2}\cos\left(2\pi\frac{f_j}{f_s}(m+a-m')\right) .
\end{aligned}
\tag{36}
$$

Thus, $\mathbf{S}^{(0)}$ is a circulant matrix, and is therefore diagonalizable in the Fourier basis:

$$
\mathbf{S}^{(0)} = \mathbf{U}\boldsymbol{\Lambda}^{(0)}\mathbf{U}^* ,
$$

where $\mathbf{U}[m,m'] = \frac{1}{\sqrt{M}} e^{-2i\pi mm'/M}$ and $\mathbf{\Lambda}^{(0)} = \text{diag}(\lambda_0^{(0)}, \ldots, \lambda_{M-1}^{(0)})$ with

$$\lambda_m^{(0)} = \sigma^2 + \sum_{j=1}^{J} \frac{\Omega_j^2}{2} \sum_{q=0}^{M-1} \cos\left(2\pi \frac{f_j}{f_s} q\right) e^{-2i\pi qm/M}$$

$$= \sigma^2 + \frac{M}{4} \sum_{j=1}^{J} \Omega_j^2 (\delta_{m,p_j} + \delta_{m,M-p_j}) .$$

Therefore,

$$\mathbf{S}^{(0)^{-1}} = \mathbf{U}\mathbf{\Lambda}^{(0)^{-1}}\mathbf{U}^* ,$$

which leads to

$$\mathbf{S}^{(0)^{-1}}[m,m'] = \frac{1}{\sigma^2}\delta_{m,m'} - \frac{2}{M\sigma^2} \sum_{j=1}^{J} \frac{1}{1 + \frac{4\sigma^2}{M\Omega_j^2}} \cos\left(2\pi p_j \frac{m-m'}{M}\right) . \tag{37}$$

Directly, we have:

$$\left|\mathbf{S}^{(0)^{-1}}[m,m']\right| = \frac{1}{\sigma^2} \left| \delta_{m,m'} - \frac{2}{M} \sum_{j=1}^{J} \frac{1}{1 + \frac{4\sigma^2}{M\Omega_j^2}} \cos\left(2\pi p_j \frac{m-m'}{M}\right) \right|$$

$$\leq \frac{1}{\sigma^2} \left( 1 + \frac{2}{M} \sum_{j=1}^{J} \frac{1}{1 + \frac{4\sigma^2}{M\Omega_j^2}} \right) \leq \frac{1}{\sigma^2}\left(1 + \frac{2J}{M}\right) .$$

Thus,

$$\left\|\mathbf{S}^{(0)^{-1}}\right\|_{\max} \leq \frac{1}{\sigma^2}\left(1 + \frac{2J}{M}\right) .$$

Furthermore, combining equations (36) and (37), we have

$$\mathbf{A}_0[m,m'] = \sum_{q=0}^{M-1} \mathbf{S}^{(1)}[m,q]\mathbf{S}^{(0)^{-1}}[q,m']$$

$$= \delta_{m+1,m'} + \frac{2}{M} \sum_{j=1}^{J} \frac{1}{1 + \frac{4\sigma^2}{M\Omega_j^2}} \cos\left(2\pi p_j \frac{m'}{M}\right) \delta_{m+1,M} .$$

Directly, we have:

$$|\mathbf{A}_0[m,m']| \leq \begin{cases} 1 & \text{if } m < M-1 , \\ \frac{2}{M}\sum_{j=1}^{J} \frac{1}{1 + \frac{4\sigma^2}{M\Omega_j^2}} & \text{if } m = M-1 . \end{cases}$$

Thus,

$$\|\mathbf{A}_0\|_{\max} \leq \max\left(1, \frac{2J}{M}\right) . \quad \blacksquare$$

**Lemma 2** (Moments of **g**). *Let* $\mathbf{g} \in \mathbb{R}^{M(M+1)}$ *be the random vector defined in* (30). *Assume the deterministic signal* $\mathbf{z}$ *takes the form* (11), *and the observed noisy signal takes the form* (13). *Then, as* $K \to \infty$, *the second-order moments of* $\mathbf{g}$ *are bounded as follows:*

$$\left|\mathbb{E}\{\mathbf{g}[r]\mathbf{g}[r']\}\right| \leq \frac{1}{K}\left(C_{\mathbf{z}}^2\sigma^2 + 2\sigma^4\right) + o\left(\frac{1}{K}\right) , \quad \forall(r,r') \in \{0,\ldots,M(M+1)-1\}^2 , \tag{38}$$

*where* $C_{\mathbf{z}} \triangleq 2\left(\sum_{j=1}^{J} \Omega_j\right)$. *Besides, higher-order moments of* $\mathbf{g}$ *behave as* $o\left(\frac{1}{K}\right)$.

*Proof.* In the following, for all $r \in \{0, \ldots, M(M+1) - 1\}$, we denote $\mathbf{g}[r] = \sigma \mathbf{g}_1[r] + \sigma^2 \mathbf{g}_2[r]$, where

$$\mathbf{g}_1[r] = \mathbf{E}_1^{(a_r)}[m_r, m_r'] = \frac{1}{K} \sum_{k=0}^{K-1} \mathbf{z}_k[m_r + a_r] \mathbf{w}_k[m_r'] + \mathbf{w}_k[m_r + a_r] \mathbf{z}_k[m_r'] ,$$

$$\mathbf{g}_2[r] = \mathbf{E}_2^{(a_r)}[m_r, m_r'] = \frac{1}{K} \sum_{k=0}^{K-1} \mathbf{w}_k[m_r + a_r] \mathbf{w}_k[m_r'] - \delta_{m_r + a_r, m_r'} ,$$

and $m_r, m_r', a_r$ are the corresponding coordinates of the matrices associated with $r$ through the vectorization operation (30). Thus, order-two moments of this random vector is split as follows:

$$\mathbf{E}\{\mathbf{g}[r]\mathbf{g}[r']\} = \sigma^2 \mathbf{E}\{\mathbf{g}_1[r]\mathbf{g}_1[r']\} + \sigma^3 \mathbf{E}\{\mathbf{g}_1[r]\mathbf{g}_2[r']\} + \sigma^3 \mathbf{E}\{\mathbf{g}_2[r]\mathbf{g}_1[r']\} + \sigma^4 \mathbf{E}\{\mathbf{g}_1[r]\mathbf{g}_1[r']\} . \quad (39)$$

By definition of the signal $\mathbf{z}$ (see equation (11)), we have $|\mathbf{z}[n]| \le \sum_{j=1}^{J} \Omega_j$, for all $n \in \mathbb{N}$. Thus, by a direct bound, we have

$$\left| \mathbf{E}\{\mathbf{g}_1[r]\mathbf{g}_1[r']\} \right| = \left| \frac{1}{K^2} \mathbf{E} \left\{ \sum_{k,k'=0}^{K-1} \mathbf{z}_k[m_r + a_r] \mathbf{z}_{k'}[m_{r'} + a_{r'}] \mathbf{w}_k[m_r'] \mathbf{w}_{k'}[m_{r'}'] + \mathbf{z}_k[m_r'] \mathbf{z}_{k'}[m_{r'} + a_{r'}] \mathbf{w}_k[m_r + a_r] \mathbf{w}_{k'}[m_{r'}'] \right. \right.$$

$$\left. \left. + \mathbf{z}_k[m_r + a_r] \mathbf{z}_{k'}[m_{r'}'] \mathbf{w}_k[m_r'] \mathbf{w}_{k'}[m_{r'} + a_{r'}] + \mathbf{z}_k[m_r'] \mathbf{z}_{k'}[m_{r'}'] \mathbf{w}_k[m_r + a_r] \mathbf{w}_{k'}[m_{r'} + a_{r'}] \right\} \right|$$

$$\le \frac{1}{K^2} \left( \sum_{j=1}^{J} \Omega_j \right)^2 \sum_{k,k'=0}^{K-1} \left| \mathbf{E}\left\{ \mathbf{w}_k[m_r'] \mathbf{w}_{k'}[m_{r'}'] \right\} \right| + \left| \mathbf{E}\left\{ \mathbf{w}_k[m_r + a_r] \mathbf{w}_{k'}[m_{r'}'] \right\} \right|$$

$$+ \left| \mathbf{E}\left\{ \mathbf{w}_k[m_r'] \mathbf{w}_{k'}[m_{r'} + a_{r'}] \right\} \right| + \left| \mathbf{E}\left\{ \mathbf{w}_k[m_r + a_r] \mathbf{w}_{k'}[m_{r'} + a_{r'}] \right\} \right| \quad (40)$$

Besides, since $\mathbf{w}$ is a white noise,

$$\frac{1}{K^2} \sum_{k,k'=0}^{K-1} \left| \mathbf{E}\left\{ \mathbf{w}_k[m_r'] \mathbf{w}_{k'}[m_{r'}'] \right\} \right| = \frac{1}{K^2} \sum_{k,k'=0}^{K-1} \delta_{k+m_r', k'+m_{r'}'}$$

$$= \frac{1}{K^2} \left( K - |m_r' - m_{r'}'| \right) .$$

Moreover, $0 \le |m_r' - m_{r'}'| \le M - 1$. Thus,

$$\frac{1}{K} - \frac{M-1}{K^2} \le \frac{1}{K^2} \sum_{k,k'=0}^{K-1} \left| \mathbf{E}\left\{ \mathbf{w}_k[m_r'] \mathbf{w}_{k'}[m_{r'}'] \right\} \right| \le \frac{1}{K} .$$

Therefore,

$$\frac{1}{K^2} \sum_{k,k'=0}^{K-1} \left| \mathbf{E}\left\{ \mathbf{w}_k[m_r'] \mathbf{w}_{k'}[m_{r'}'] \right\} \right| = \frac{1}{K} + o\left( \frac{1}{K} \right) .$$

Similar calculations lead to the same results for the other three terms making up the sum (40). Therefore, we have:

$$\left| \mathbf{E}\{\mathbf{g}_1[r]\mathbf{g}_1[r']\} \right| \le \left( \sum_{j=1}^{J} \Omega_j \right)^2 \left( \frac{4}{K} + o\left( \frac{1}{K} \right) \right)$$

$$\le \frac{C_{\mathbf{z}}^2}{K} + o\left( \frac{1}{K} \right) . \quad (41)$$

Besides, since odd-order moments of a zero-mean multivariate Gaussian random vector are zero, we have:

$$\mathbf{E}\{\mathbf{g}_1[r]\mathbf{g}_2[r']\} = \frac{1}{K^2} \sum_{k,k'=0}^{K-1} \mathbf{z}_k[m_r + a_r] \mathbf{E}\left\{ \mathbf{w}_k[m_r'] \mathbf{w}_{k'}[m_{r'} + a_{r'}] \mathbf{w}_{k'}[m_{r'}'] \right\} + \mathbf{z}_k[m_r'] \mathbf{E}\left\{ \mathbf{w}_k[m_r + a_r] \mathbf{w}_{k'}[m_{r'} + a_{r'}] \mathbf{w}_{k'}[m_{r'}'] \right\}$$

$$- \delta_{m_{r'} + a_{r'}, m_{r'}'} \mathbf{E}\{\mathbf{g}_1[r]\}$$

$$= 0 . \quad (42)$$

Similarly,

$$\mathbf{E}\{\mathbf{g}_2[r]\mathbf{g}_1[r']\} = 0 . \quad (43)$$

Besides, by a direct calculation, we have:

$$\mathbf{E}\{\mathbf{g}_2[r]\mathbf{g}_2[r']\} = \frac{1}{K^2}\sum_{k,k'=0}^{K-1}\mathbb{E}\left\{\mathbf{w}_k[m_r+a_r]\mathbf{w}_k[m_r']\mathbf{w}_{k'}[m_{r'}+a_{r'}]\mathbf{w}_{k'}[m_{r'}']\right\} - \delta_{m_r+a_r,m_r'}\frac{1}{K}\sum_{k'=0}^{K-1}\mathbb{E}\left\{\mathbf{w}_{k'}[m_{r'}+a_{r'}]\mathbf{w}_{k'}[m_{r'}']\right\}$$

$$- \delta_{m_{r'}+a_{r'},m_{r'}'}\frac{1}{K}\sum_{k=0}^{K-1}\mathbb{E}\left\{\mathbf{w}_k[m_r']\mathbf{w}_{k'}[m_{r'}']\right\} + \delta_{m_r+a_r,m_r'}\delta_{m_{r'}+a_{r'},m_{r'}'}$$

$$= \frac{1}{K^2}\sum_{k,k'=0}^{K-1}\mathbb{E}\left\{\mathbf{w}_k[m_r+a_r]\mathbf{w}_k[m_r']\mathbf{w}_{k'}[m_{r'}+a_{r'}]\mathbf{w}_{k'}[m_{r'}']\right\} - \delta_{m_r+a_r,m_r'}\delta_{m_{r'}+a_{r'},m_{r'}'} \tag{44}$$

Moreover, using the results of the Isserlis' theorem [1] to express fourth-order moments of a Gaussian random vector as a function of its second-order moments, we expand sum (44) as follows:

$$\mathbf{E}\{\mathbf{g}_2[r]\mathbf{g}_2[r']\} = \frac{1}{K^2}\sum_{k,k'=0}^{K-1}\mathbb{E}\left\{\mathbf{w}_k[m_r+a_r]\mathbf{w}_k[m_r']\right\}\mathbb{E}\left\{\mathbf{w}_{k'}[m_{r'}+a_{r'}]\mathbf{w}_{k'}[m_{r'}']\right\} - \delta_{m_r+a_r,m_r'}\delta_{m_{r'}+a_{r'},m_{r'}'}$$

$$+ \frac{1}{K^2}\sum_{k,k'=0}^{K-1}\mathbb{E}\left\{\mathbf{w}_k[m_r+a_r]\mathbf{w}_{k'}[m_{r'}+a_{r'}]\right\}\mathbb{E}\left\{\mathbf{w}_k[m_r']\mathbf{w}_{k'}[m_{r'}']\right\}$$

$$+ \frac{1}{K^2}\sum_{k,k'=0}^{K-1}\mathbb{E}\left\{\mathbf{w}_k[m_r+a_r]\mathbf{w}_{k'}[m_{r'}']\right\}\mathbb{E}\left\{\mathbf{w}_k[m_r']\mathbf{w}_{k'}[m_{r'}+a_{r'}]\right\}$$

$$= \frac{1}{K^2}\sum_{k,k'=0}^{K-1}\delta_{k+m_r+a_r,k'+m_{r'}+a_{r'}}\delta_{k+m'r,k'+m_{r'}'} + \delta_{k+m_r+a_r,k'+m_{r'}'}\delta_{k+m'r,k'+m_{r'}+a_{r'}}$$

$$= \frac{1}{K^2}\left(\delta_{m_r'-m_{r'}',m_r+a_r-m_{r'}-a_{r'}}(K-|m_r'-m_{r'}'|) + \delta_{m_r+a_r-m_{r'}',m_r'-m_{r'}-a_{r'}}(K-|m_r+a_r-m_{r'}'|)\right)$$

$$= \frac{1}{K}\left(\delta_{m_r'-m_{r'}',m_r+a_r-m_{r'}-a_{r'}} + \delta_{m_r'+m_{r'}',m_r+a_r+m_{r'}+a_{r'}}\right) + o\left(\frac{1}{K}\right).$$

Therefore,

$$\left|\mathbf{E}\{\mathbf{g}_2[r]\mathbf{g}_2[r']\}\right| \le \frac{2}{K} + o\left(\frac{1}{K}\right). \tag{45}$$

Thus, combining results (41), (42), (43) and (45) into expression (39) gives the following bound:

$$\left|\mathbf{E}\{\mathbf{g}[r]\mathbf{g}[r']\}\right| \le \frac{1}{K}\left(C_\mathbf{z}^2\sigma^2 + 2\sigma^4\right) + o\left(\frac{1}{K}\right).$$

Concerning higher-order moments, let $T \ge 3$ denote the order of the moment defined by $\mathbb{E}\{\prod_{\theta=1}^T \mathbf{g}[r_\theta]\}$. Thus,

$$\mathbb{E}\left\{\prod_{\theta=1}^T \mathbf{g}[r_\theta]\right\} = \mathbb{E}\left\{\prod_{\theta=1}^T\left(\sigma\mathbf{g}_1[r_\theta] + \sigma^2\mathbf{g}_2[r_\theta]\right)\right\}$$

$$= \frac{1}{K^T}\mathbb{E}\left\{\prod_{\theta=1}^T\left(\sum_{k=0}^{K-1}\rho_{\theta,k}\left(\mathbf{w}[k+m_{r_\theta}'],\mathbf{w}[k+m_{r_\theta}+a_{r_\theta}]\right)\right)\right\}$$

$$= \frac{1}{K^T}\sum_{k_1=0}^{K-1}\cdots\sum_{k_T=0}^{K-1}\mathbb{E}\left\{\prod_{\theta=1}^T\rho_{\theta,k_\theta}\left(\mathbf{w}[k_\theta+m_{r_\theta}'],\mathbf{w}[k_\theta+m_{r_\theta}+a_{r_\theta}]\right)\right\}, \tag{46}$$

where

$$\rho_{\theta,k}(u,v) = \sigma\mathbf{z}_k[m_{r_\theta}+a_{r_\theta}]u + \sigma\mathbf{z}_k[m_{r_\theta}']v + \sigma^2 uv - \delta_{m_{r_\theta}+a_{r_\theta},m_{r_\theta}'}.$$

Thus,

$$\left|\mathbb{E}\left\{\prod_{\theta=1}^T \mathbf{g}[r_\theta]\right\}\right| \le \frac{\nu_K}{K^T}C_T,$$

where

$$C_T = \max_{(k_1,\ldots,k_T)}\left|\mathbb{E}\left\{\prod_{\theta=1}^T\rho_{\theta,k_\theta}\left(\mathbf{w}[k_\theta+m_{r_\theta}'],\mathbf{w}[k_\theta+m_{r_\theta}+a_{r_\theta}]\right)\right\}\right|,$$

and $\nu_K$ is the number of nonzero terms in the sum (46). Note that $C_T$ is independent of $K$ (but depends on $\mathbf{z}$ and $\sigma$). The behavior of the order $T$ moment in function of $K$ is therefore only determined by the ratio $\frac{\nu_K}{K^T}$.

Let us bound $\nu_K$. Fix $k_1$. For each of the other indexes of summation $k_\theta$ ($\theta \in \{2, \ldots, T\}$), there are four values that make $\rho_{\theta,k_\theta}(\mathbf{w}[k_\theta + m'_{r_\theta}], \mathbf{w}[k_\theta + m_{r_\theta} + a_{r_\theta}])$ not independent of $\rho_{1,k_1}(\mathbf{w}[k_1 + m'_{r_1}], \mathbf{w}[k_1 + m_{r_1} + a_{r_1}])$. Indeed, since $\mathbf{w}$ is a white noise these quantities are independent except when $k_\theta$ is such that:

$$k_\theta = k_1 + m'_{r_1} - m'_{r_\theta}$$
$$k_\theta = k_1 + m_{r_1} + a_{r_1} - m'_{r_\theta}$$
$$k_\theta = k_1 + m'_{r_1} - m_{r_\theta} - a_{r_\theta}$$
$$k_\theta = k_1 + m_{r_1} + a_{r_1} - m_{r_\theta} - a_{r_\theta} \ .$$

Consequently, for each value of $k_1$ there exist at least $(K-4)^{T-1}$ combinations of $(k_2, \ldots, k_T)$ where we have

$$\mathbb{E}\left\{\prod_{\theta=1}^{T} \rho_{\theta,k_\theta}(\mathbf{w}[k_\theta + m'_{r_\theta}], \mathbf{w}[k_\theta + m_{r_\theta} + a_{r_\theta}])\right\} = \mathbb{E}\left\{\rho_{1,k_1}(\mathbf{w}[k_1 + m'_{r_1}], \mathbf{w}[k_1 + m_{r_1} + a_{r_1}])\right\}$$

$$\times \mathbb{E}\left\{\prod_{\theta=2}^{T} \rho_{\theta,k_\theta}(\mathbf{w}[k_\theta + m'_{r_\theta}], \mathbf{w}[k_\theta + m_{r_\theta} + a_{r_\theta}])\right\} = 0,$$

because $\mathbb{E}\left\{\rho_{1,k_1}(\mathbf{w}[k_1 + m'_{r_1}], \mathbf{w}[k_1 + m_{r_1} + a_{r_1}])\right\} = 0$. Therefore, at least $K(K-4)^{T-1}$ of the sum (46) are zero. Because $T \geq 3$, we develop similar arguments on $k_2$ and $k_3$ to determine other cases where this correlation term vanishes. We subtract these cases to $K^T$, the total number of combinations of $(k_1, \ldots, k_T)$ to obtain the following maximum bound on the number of nonzero terms in the sum (46):

$$\nu_K \leq K^T - 3K(K-4)^{T-1} + 3K(K-4)(K-8)^{T-2} - K(K-4)(K-8)(K-12)^{T-3} \ .$$

Thus,

$$\left|\mathbb{E}\left\{\prod_{\theta=1}^{T} \mathbf{g}[r_\theta]\right\}\right| \leq C_T \frac{K^T - 3K(K-4)^{T-1} + 3K(K-4)(K-8)^{T-2} - K(K-4)(K-8)(K-12)^{T-3}}{K^T}$$

$$\leq C_T\left(1 - 3\left(1 - \frac{4}{K}\right)^{T-1} + 3\left(1 - \frac{4}{K}\right)\left(1 - \frac{8}{K}\right)^{T-2} - \left(1 - \frac{4}{K}\right)\left(1 - \frac{8}{K}\right)\left(1 - \frac{12}{K}\right)^{T-3}\right)$$

$$\leq C_T\left(1 - 3 + \frac{12(T-1)}{K} + 3 - \frac{12}{K} - \frac{24(T-2)}{K} - 1 + \frac{4}{K} + \frac{8}{K} + \frac{12(T-3)}{K} + o\left(\frac{1}{K}\right)\right)$$

$$\leq C_T \, o\left(\frac{1}{K}\right) \ .$$

Therefore,

$$\left|\mathbb{E}\left\{\prod_{\theta=1}^{T} \mathbf{g}[r_\theta]\right\}\right| = o\left(\frac{1}{K}\right), \quad \forall T \geq 3 \ . \ \blacksquare$$

**Lemma 3** (Bounds on the derivatives of $f^{(\ell)}$ at the origin). *Let $f^{(\ell)} : \mathbb{R}^{M(M+1)} \to \mathbb{R}^M$ denote the multivariate function defined in (31). Assume the deterministic signal $\mathbf{z}$ takes the form (11), and the observed noisy signal takes the form (13). Then, the first-order derivatives of $f^{(\ell)}$ at the origin are bounded as follows:*

$$\left|\frac{\partial f_m^{(\ell)}}{\partial \mathbf{g}[r]}\bigg|_{\mathbf{g}=0}\right| \leq \frac{d_{1,\mathbf{z},M,\ell}}{\sigma^2}, \quad \forall m \in \{0, \ldots, M-1\}, \ r \in \{0, \ldots, M(M+1)-1\} \ , \tag{47}$$

*where*

$$d_{1,\mathbf{z},M,\ell} \triangleq (2 + (\ell-2)M) \, M^{\ell-1}\left(\max\left(1, \frac{2J}{M}\right)\right)^{\ell-1}\left(1 + \max\left(1, \frac{2J}{M}\right)\right)\left(1 + \frac{2J}{M}\right) \ .$$

*Besides, the second-order derivatives of $f^{(\ell)}$ at the origin are bounded as follows:*

$$\left|\frac{\partial^2 f_m^{(\ell)}}{\partial \mathbf{g}[r]\partial \mathbf{g}[r']}\bigg|_{\mathbf{g}=0}\right| \leq \frac{d_{2,\mathbf{z},M,\ell}}{\sigma^4}, \quad \forall m \in \{0, \ldots, M-1\}, \ (r,r') \in \{0, \ldots, M(M+1)-1\}^2 \ , \tag{48}$$

*where*

$$d_{2,\mathbf{z},M,\ell} \triangleq \left(1 + \frac{2J}{M}\right)^2\left(\max\left(1, \frac{2J}{M}\right)\right)^{\ell-2}\left(1 + \max\left(1, \frac{2J}{M}\right)\right)\left(d_{2,M,\ell} + (d_{2,M,\ell} + 2d'_{2,M,\ell})\max\left(1, \frac{2J}{M}\right)\right) \ ,$$

and $d_{2,M,\ell}$ and $d'_{2,M,\ell}$ are only depending on $M$ and $\ell$.

*Proof.* Concerning the first-order derivative, from (29), we have:

$$\frac{\partial f^{(\ell)}}{\partial \mathbf{g}[r]} = \mathbf{e}_M^T \frac{\partial \tilde{\mathbf{A}}^\ell}{\partial \mathbf{g}[r]}$$

$$= \sum_{\lambda=0}^{\ell-1} \mathbf{e}_M^T \tilde{\mathbf{A}}^\lambda \frac{\partial \tilde{\mathbf{A}}}{\partial \mathbf{g}[r]} \tilde{\mathbf{A}}^{\ell-1-\lambda} .$$

Thus,

$$\left. \frac{\partial f^{(\ell)}}{\partial \mathbf{g}[r]} \right|_{\mathbf{g}=0} = \sum_{\lambda=0}^{\ell-1} \mathbf{e}_M^T \mathbf{A}_0^\lambda \left. \frac{\partial \tilde{\mathbf{A}}}{\partial \mathbf{g}[r]} \right|_{\mathbf{g}=0} \mathbf{A}_0^{\ell-1-\lambda} . \tag{49}$$

Furthermore,

$$\frac{\partial \tilde{\mathbf{A}}}{\partial \mathbf{g}[r]} = \frac{\partial \mathbf{E}^{(1)}}{\partial \mathbf{g}[r]} \left( \mathbf{S}^{(0)} + \mathbf{E}^{(0)} \right)^{-1} + \left( \mathbf{S}^{(1)} + \mathbf{E}^{(1)} \right) \frac{\partial \left( \mathbf{S}^{(0)} + \mathbf{E}^{(0)} \right)^{-1}}{\partial \mathbf{g}[r]}$$

$$= \begin{cases} \mathbf{J}_{m_r,m_r'} \left( \mathbf{S}^{(0)} + \mathbf{E}^{(0)} \right)^{-1} & \text{if } a_r = 1, \\ \left( (1-\delta_{m_r,0}) \mathbf{J}_{m_r-1,m_r'} + \left( \mathbf{S}^{(1)} + \mathbf{E}^{(1)} \right) \left( \mathbf{S}^{(0)} + \mathbf{E}^{(0)} \right)^{-1} \mathbf{J}_{m_r,m_r'} \right) \left( \mathbf{S}^{(0)} + \mathbf{E}^{(0)} \right)^{-1} & \text{else,} \end{cases}$$

where $\mathbf{J}_{m_r,m_r'} \in \mathbb{R}^{M\times M}$ is the matrix such that $\mathbf{J}_{m_r,m_r'}[m,m'] \triangleq \delta_{m,m_r} \delta_{m',m_r'}$. Thus,

$$\left. \frac{\partial \tilde{\mathbf{A}}}{\partial \mathbf{g}[r]} \right|_{\mathbf{g}=0} = \begin{cases} \mathbf{J}_{m_r,m_r'} \mathbf{S}^{(0)^{-1}} & \text{if } a_r = 1, \\ \left( (1-\delta_{m_r,0}) \mathbf{J}_{m_r-1,m_r'} + \mathbf{A}_0 \mathbf{J}_{m_r,m_r'} \right) \mathbf{S}^{(0)^{-1}} & \text{else.} \end{cases}$$

Then,

$$\left\| \left. \frac{\partial \tilde{\mathbf{A}}}{\partial \mathbf{g}[r]} \right|_{\mathbf{g}=0} \right\|_{\max} \leq \left( 1 + \|\mathbf{A}_0\|_{\max} \right) \left\| \mathbf{S}^{(0)^{-1}} \right\|_{\max} .$$

Given bounds (35) and (34), we have that:

$$\left\| \left. \frac{\partial \tilde{\mathbf{A}}}{\partial \mathbf{g}[r]} \right|_{\mathbf{g}=0} \right\|_{\max} \leq \left( 1 + \max \left( 1, \frac{2J}{M} \right) \right) \left( 1 + \frac{2J}{M} \right) \frac{1}{\sigma^2} . \tag{50}$$

Besides given expression (49), for all $r \in \{0, \dots, M(M+1)-1\}$, we have:

$$\left| \left. \frac{\partial f_m^{(\ell)}}{\partial \mathbf{g}[r]} \right|_{\mathbf{g}=0} \right| \leq 2M \|\mathbf{A}_0^{\ell-1}\|_{\max} \left\| \left. \frac{\partial \tilde{\mathbf{A}}}{\partial \mathbf{g}[r]} \right|_{\mathbf{g}=0} \right\|_{\max} + M^2 \sum_{\lambda=1}^{\ell-2} \|\mathbf{A}_0^\lambda\|_{\max} \left\| \left. \frac{\partial \tilde{\mathbf{A}}}{\partial \mathbf{g}[r]} \right|_{\mathbf{g}=0} \right\|_{\max} \|\mathbf{A}_0^{\ell-1-\lambda}\|_{\max}$$

$$\leq \left( 2 + (\ell-2)M \right) M^{\ell-1} \|\mathbf{A}_0\|_{\max}^{\ell-1} \left\| \left. \frac{\partial \tilde{\mathbf{A}}}{\partial \mathbf{g}[r]} \right|_{\mathbf{g}=0} \right\|_{\max} .$$

Therefore, given bounds (35) and (50), we have:

$$\left| \left. \frac{\partial f_m^{(\ell)}}{\partial \mathbf{g}[r]} \right|_{\mathbf{g}=0} \right| \leq \frac{d_{1,\mathbf{z},M,\ell}}{\sigma^2} .$$

Concerning the second-order derivative, we have:

$$\frac{\partial^2 f^{(\ell)}}{\partial \mathbf{g}[r] \partial \mathbf{g}[r']} = \sum_{\lambda=0}^{\ell-1} \mathbf{e}_M^T \frac{\partial \tilde{\mathbf{A}}^\lambda}{\partial \mathbf{g}[r']} \frac{\partial \tilde{\mathbf{A}}}{\partial \mathbf{g}[r]} \tilde{\mathbf{A}}^{\ell-1-\lambda} + \mathbf{e}_M^T \tilde{\mathbf{A}}^\lambda \frac{\partial^2 \tilde{\mathbf{A}}}{\partial \mathbf{g}[r] \partial \mathbf{g}[r']} \tilde{\mathbf{A}}^{\ell-1-\lambda} + \mathbf{e}_M^T \tilde{\mathbf{A}}^\lambda \frac{\partial \tilde{\mathbf{A}}}{\partial \mathbf{g}[r]} \frac{\partial \tilde{\mathbf{A}}^{\ell-1-\lambda}}{\partial \mathbf{g}[r']}$$

$$= \sum_{\lambda=1}^{\ell-1} \sum_{p=0}^{\lambda-1} \mathbf{e}_M^T \tilde{\mathbf{A}}^p \frac{\partial \tilde{\mathbf{A}}}{\partial \mathbf{g}[r']} \tilde{\mathbf{A}}^{\lambda-1-p} \frac{\partial \tilde{\mathbf{A}}}{\partial \mathbf{g}[r]} \tilde{\mathbf{A}}^{\ell-1-\lambda} + \sum_{\lambda=0}^{\ell-1} \mathbf{e}_M^T \tilde{\mathbf{A}}^l \frac{\partial^2 \tilde{\mathbf{A}}}{\partial \mathbf{g}[r] \partial \mathbf{g}[r']} \tilde{\mathbf{A}}^{\ell-1-\lambda}$$

$$+ \sum_{\lambda=0}^{\ell-2} \sum_{p=0}^{\ell-\lambda-2} \mathbf{e}_M^T \tilde{\mathbf{A}}^\lambda \frac{\partial \tilde{\mathbf{A}}}{\partial \mathbf{g}[r]} \tilde{\mathbf{A}}^p \frac{\partial \tilde{\mathbf{A}}}{\partial \mathbf{g}[r']} \tilde{\mathbf{A}}^{\ell-\lambda-2-p} .$$

Thus,

$$\left.\frac{\partial^2 f^{(\ell)}}{\partial \mathbf{g}[r]\partial \mathbf{g}[r']}\right|_{\mathbf{g}=0} = \sum_{\lambda=1}^{\ell-1}\sum_{p=0}^{\lambda-1} \mathbf{e}_M^T \mathbf{A}_0^p \left.\frac{\partial \tilde{\mathbf{A}}}{\partial \mathbf{g}[r']}\right|_{\mathbf{g}=0} \mathbf{A}_0^{\lambda-1-p} \left.\frac{\partial \tilde{\mathbf{A}}}{\partial \mathbf{g}[r]}\right|_{\mathbf{g}=0} \mathbf{A}_0^{\ell-1-\lambda}$$

$$+ \sum_{\lambda=0}^{\ell-1} \mathbf{e}_M^T \mathbf{A}_0^\lambda \left.\frac{\partial^2 \tilde{\mathbf{A}}}{\partial \mathbf{g}[r]\partial \mathbf{g}[r']}\right|_{\mathbf{g}=0} \mathbf{A}_0^{\ell-1-\lambda} + \sum_{\lambda=0}^{\ell-2}\sum_{p=0}^{\ell-2-\lambda} \mathbf{e}_M^T \mathbf{A}_0^\lambda \left.\frac{\partial \tilde{\mathbf{A}}}{\partial \mathbf{g}[r]}\right|_{\mathbf{g}=0} \mathbf{A}_0^p \left.\frac{\partial \tilde{\mathbf{A}}}{\partial \mathbf{g}[r']}\right|_{\mathbf{g}=0} \mathbf{A}_0^{\ell-\lambda-2-p} . \quad (51)$$

Besides, the second-order derivative of the matrix $\tilde{\mathbf{A}}$ is given by

$$\frac{\partial^2 \tilde{\mathbf{A}}}{\partial \mathbf{g}[r]\partial \mathbf{g}[r']} = \begin{cases} 0 & \text{if } a_r = 1 \text{ and } a_r' = 1, \\ \mathbf{J}_{m_r,m_r'}\left(\mathbf{S}^{(0)}+\mathbf{E}^{(0)}\right)^{-1}\mathbf{J}_{m_{r'},m_{r'}'}\left(\mathbf{S}^{(0)}+\mathbf{E}^{(0)}\right)^{-1} & \text{if } a_r = 1 \text{ and } a_r' = 0, \\ \mathbf{J}_{m_{r'}',m_{r'}'}\left(\mathbf{S}^{(0)}+\mathbf{E}^{(0)}\right)^{-1}\mathbf{J}_{m_r,m_r'}\left(\mathbf{S}^{(0)}+\mathbf{E}^{(0)}\right)^{-1} & \text{if } a_r = 0 \text{ and } a_r' = 1, \\ \left((1-\delta_{m_r,0})\mathbf{J}_{m_r-1,m_r'}+\left(\mathbf{S}^{(1)}+\mathbf{E}^{(1)}\right)\left(\mathbf{S}^{(0)}+\mathbf{E}^{(0)}\right)^{-1}\mathbf{J}_{m_r,m_r'}\right) & \\ \quad \times \left(\mathbf{S}^{(0)}+\mathbf{E}^{(0)}\right)^{-1}\mathbf{J}_{m_{r'},m_{r'}'}\left(\mathbf{S}^{(0)}+\mathbf{E}^{(0)}\right)^{-1} & \\ \quad + \left((1-\delta_{m_{r'},0})\mathbf{J}_{m_{r'}-1,m_{r'}'}+\left(\mathbf{S}^{(1)}+\mathbf{E}^{(1)}\right)\left(\mathbf{S}^{(0)}+\mathbf{E}^{(0)}\right)^{-1}\mathbf{J}_{m_{r'},m_{r'}'}\right) & \\ \quad \times \left(\mathbf{S}^{(0)}+\mathbf{E}^{(0)}\right)^{-1}\mathbf{J}_{m_r,m_r'}\left(\mathbf{S}^{(0)}+\mathbf{E}^{(0)}\right)^{-1} & \text{else.} \end{cases}$$

Thus, the second-order derivative of the matrix $\tilde{\mathbf{A}}$ at the origin is such that

$$\left.\frac{\partial^2 \tilde{\mathbf{A}}}{\partial \mathbf{g}[r]\partial \mathbf{g}[r']}\right|_{\mathbf{g}=0} = \begin{cases} 0 & \text{if } a_r = 1 \text{ and } a_{r'} = 1, \\ \mathbf{J}_{m_r,m_r'}\mathbf{S}^{(0)^{-1}}\mathbf{J}_{m_{r'},m_{r'}'}\mathbf{S}^{(0)^{-1}} & \text{if } a_r = 1 \text{ and } a_{r'} = 0, \\ \mathbf{J}_{m_{r'},m_{r'}'}\mathbf{S}^{(0)^{-1}}\mathbf{J}_{m_r,m_r'}\mathbf{S}^{(0)^{-1}} & \text{if } a_r = 0 \text{ and } a_{r'} = 1, \\ \left((1-\delta_{m_r,0})\mathbf{J}_{m_r-1,m_r'}+\mathbf{A}_0\mathbf{J}_{m_r,m_r'}\right)\mathbf{S}^{(0)^{-1}}\mathbf{J}_{m_{r'},m_{r'}'}\mathbf{S}^{(0)^{-1}} & \\ \quad + \left((1-\delta_{m_{r'},0})\mathbf{J}_{m_{r'}-1,m_{r'}'}+\mathbf{A}_0\mathbf{J}_{m_{r'},m_{r'}'}\right)\mathbf{S}^{(0)^{-1}}\mathbf{J}_{m_r,m_r'}\mathbf{S}^{(0)^{-1}} & \text{else.} \end{cases}$$

Then,

$$\left\|\left.\frac{\partial^2 \tilde{\mathbf{A}}}{\partial \mathbf{g}[r]\partial \mathbf{g}[r']}\right|_{\mathbf{g}=0}\right\|_{\max} \leq \begin{cases} \left\|\mathbf{S}^{(0)^{-1}}\right\|_{\max}^2 & \text{if } a_r = 1 \text{ or } a_{r'} = 1, \\ 2\left(1+\|\mathbf{A}_0\|_{\max}\right)\left\|\mathbf{S}^{(0)^{-1}}\right\|_{\max}^2 & \text{else} \end{cases}$$

$$\leq 2\left(1+\max\left(1,\frac{2J}{M}\right)\right)\left(1+\frac{2J}{M}\right)^2 \frac{1}{\sigma^4} . \quad (52)$$

Returning to equation (51), for all $r, r' \in \{0,\ldots,M(M+1)-1\}$ and $m \in \{0,\ldots,M-1\}$, we have:

$$\left|\left.\frac{\partial^2 f_m^{(\ell)}}{\partial \mathbf{g}[r]\partial \mathbf{g}[r']}\right|_{\mathbf{g}=0}\right| \leq d_{2,M,\ell}\|\mathbf{A}_0\|_{\max}^{\ell-2}\left\|\frac{\partial \tilde{\mathbf{A}}}{\partial \mathbf{g}[r]}\right\|_{\max}\left\|\frac{\partial \tilde{\mathbf{A}}}{\partial \mathbf{g}[r']}\right\|_{\max} + d_{2,M,\ell}'\|\mathbf{A}_0\|_{\max}^{\ell-1}\left\|\frac{\partial^2 \tilde{\mathbf{A}}}{\partial \mathbf{g}[r]\mathbf{g}[r']}\right\|_{\max} ,$$

where $d_{2,M,\ell}$ and $d_{2,M,\ell}'$ are only depending on $M$ and $\ell$. Besides, given results (28), (50) and (52), we have:

$$\left|\left.\frac{\partial^2 f_m^{(\ell)}}{\partial \mathbf{g}[r]\partial \mathbf{g}[r']}\right|_{\mathbf{g}=0}\right| \leq \frac{d_{2,\mathbf{z},M,\ell}}{\sigma^4} . \quad \blacksquare$$

### B. Expression of the Bias $\boldsymbol{\mu}$.

By definition of the measurement noise, $\boldsymbol{\mu}[n] = 0$ when $n \in I$. Outside the measurement interval $I$, denote by $\ell$ the index such that $n = N-1+\ell$. Then, given that $\mathbf{h}^{(\ell)} = \boldsymbol{\alpha}^{(\ell)} - \boldsymbol{\alpha}_0^{(\ell)}$, we deduce from expression (16) that

$$\boldsymbol{\mu}[n] = \mathbb{E}\{\boldsymbol{\alpha}^{(\ell)}\}\mathbf{z}_K + \sigma\mathbb{E}\{\boldsymbol{\alpha}^{(\ell)}\mathbf{w}_K\} - \mathbf{z}[n]$$

$$= \boldsymbol{\alpha}_0^{(\ell)}\mathbf{z}_K + \mathbb{E}\{\mathbf{h}^{(\ell)}\}\mathbf{z}_K + \sigma\mathbb{E}\{\mathbf{h}^{(\ell)}\mathbf{w}_K\} - \mathbf{z}[N-1+\ell]$$

$$\stackrel{\Delta}{=} \epsilon_1[\ell] + \epsilon_2[\ell] + \epsilon_3[\ell] , \quad (53)$$

where

$$\epsilon_1[\ell] = \boldsymbol{\alpha}_0^{(\ell)} \mathbf{z}_K - \mathbf{z}[N-1+\ell] , \tag{54}$$

$$\epsilon_2[\ell] = \mathbb{E}\{\mathbf{h}^{(\ell)}\}\mathbf{z}_K , \tag{55}$$

$$\epsilon_3[\ell] = \sigma \mathbb{E}\{\mathbf{h}^{(\ell)}\mathbf{w}_K\} . \tag{56}$$

Let us first determine an upper bound on $|\epsilon_1[n]|$. Since $\boldsymbol{\alpha}_0^{(1)}$ is the last row of $\mathbf{A}_0$, we deduce from the expression (33) of $\mathbf{A}_0$ that

$$\boldsymbol{\alpha}_0^{(1)}[m] = \frac{2}{M} \sum_{j=1}^{J} \frac{1}{1 + \frac{4\sigma^2}{M\Omega_j^2}} \cos\left(2\pi p_j \frac{m}{M}\right) . \tag{57}$$

Besides, from equation (33), we also have

$$\mathbf{A}_0 \mathbf{z}_K = \begin{pmatrix} \mathbf{z}[N-M+1] \\ \vdots \\ \mathbf{z}[N-1] \\ \boldsymbol{\alpha}_0^{(1)} \mathbf{z}_K \end{pmatrix} .$$

The upward-shift property is thus successively inducted when $\ell$ increases; that is,

$$\mathbf{A}_0^{\ell} \mathbf{z}_K = \begin{pmatrix} \mathbf{z}[N-M+\ell] \\ \vdots \\ \mathbf{z}[N-1] \\ \boldsymbol{\alpha}_0^{(1)} \mathbf{z}_K \\ \vdots \\ \boldsymbol{\alpha}_0^{(\ell)} \mathbf{z}_K \end{pmatrix} .$$

Then, $\boldsymbol{\alpha}_0^{(\ell)}$, the last row of $\mathbf{A}_0^{\ell}$ follows the following recurrence relation:

$$\boldsymbol{\alpha}_0^{(\ell)} \mathbf{z}_K = \boldsymbol{\alpha}_0^{(1)} \tilde{\mathbf{A}}_0^{\ell-1} \mathbf{z}_K$$
$$= \sum_{m=0}^{M-\ell} \boldsymbol{\alpha}_0^{(1)}[m] \mathbf{z}[N-M+\ell+m-1] + \sum_{m=M-\ell+1}^{M-1} \boldsymbol{\alpha}_0^{(1)}[m] \boldsymbol{\alpha}_0^{(m-M+\ell)} \mathbf{z}_K .$$

Hence,

$$\epsilon_1[\ell] = \boldsymbol{\alpha}_0^{(\ell)} \mathbf{z}_K - \mathbf{z}[N-1+\ell]$$
$$= \sum_{m=0}^{M} \boldsymbol{\alpha}_0^{(1)}[m] \mathbf{z}[N-M+\ell+m-1] - \mathbf{z}[N-1+\ell] + \sum_{m=M-\ell+1}^{M-1} \boldsymbol{\alpha}_0^{(1)}[m] \left( \boldsymbol{\alpha}_0^{(m-M+\ell)} \mathbf{z}_K - \mathbf{z}[N-M+\ell+m-1] \right)$$
$$= \sum_{m=0}^{M} \boldsymbol{\alpha}_0^{(1)}[m] \mathbf{z}[N-M+\ell+m-1] - \mathbf{z}[N-1+\ell] + \sum_{m=M-\ell+1}^{M-1} \boldsymbol{\alpha}_0^{(1)}[m] \epsilon_1[m-M+\ell] .$$

Besides, equation (57) gives

$$\sum_{m=0}^{M} \boldsymbol{\alpha}_0^{(1)}[m] \mathbf{z}[N-M+\ell+m-1] = \sum_{j,j'=1}^{J} \Omega_{j'} \frac{2}{M} \frac{1}{1+\frac{4\sigma^2}{M\Omega_j^2}} \underbrace{\sum_{m=0}^{M-1} \cos\left(2\pi p_j \frac{m}{M}\right) \cos\left(2\pi p_{j'} \frac{N+m}{M} + \varphi_{j'}\right)}_{=\delta_{j,j'} \frac{M}{2} \cos\left(2\pi p_j \frac{N}{M} + \varphi_j\right)}$$
$$= \sum_{j=1}^{J} \Omega_j \frac{1}{1+\frac{4\sigma^2}{M\Omega_j^2}} \cos\left(2\pi p_j \frac{N}{M} + \varphi_j\right) .$$

Thus:

$$|\epsilon_1[\ell]| = \left| \sum_{j=1}^{J} \Omega_j \left( \frac{1}{1 + \frac{4\sigma^2}{M\Omega_j^2}} - 1 \right) \cos\left( 2\pi p_j \frac{N}{M} + \varphi_j \right) + \sum_{m=M-\ell+1}^{M-1} \boldsymbol{\alpha}_0^{(1)}[m] \epsilon_1[m - M + \ell] \right|$$

$$\leq \sum_{j=1}^{J} \Omega_j \left| \frac{1}{1 + \frac{4\sigma^2}{M\Omega_j^2}} - 1 \right| + \frac{2J}{M} \sum_{\lambda=1}^{\ell-1} |\epsilon_1[\lambda]|$$

$$\leq \frac{4\sigma^2}{M} \sum_{j=1}^{J} \frac{1}{\Omega_j} + \frac{2J}{M} \sum_{\lambda=1}^{\ell-1} |\epsilon_1[\lambda]| \; . \tag{58}$$

Then, by induction from the inequality (58), we have

$$|\epsilon_1[\ell]| \leq \frac{4\sigma^2}{M} \left( 1 + \frac{2J}{M} \right)^{\ell-1} \sum_{j=1}^{J} \frac{1}{\Omega_j} \triangleq c^{(1)} \sigma^2 \; , \tag{59}$$

where $c^{(1)} = \frac{4}{M} \left( 1 + \frac{2J}{M} \right)^{\ell-1} \sum_{j=1}^{J} \frac{1}{\Omega_j}$. Note that $c^{(1)}$ is not depending on $\sigma$ or $K$.

Let us now determine an upper bound on $|\epsilon_2[\ell]|$. Since $\mathbb{E}\{\mathbf{g}[r]\} = 0$ and moments of order 3 and higher behave as $o\left( \frac{1}{K} \right)$, a second-order Taylor expansion of $\mathbf{h}^{(\ell)}$ gives

$$\mathbb{E}\{\mathbf{h}^{(\ell)}[m]\} = \frac{1}{2} \sum_{r,r'=0}^{M(M+1)-1} \left. \frac{\partial^2 f_m^{(\ell)}}{\partial \mathbf{g}[r] \partial \mathbf{g}[r']} \right|_{\mathbf{g}=0} \mathbb{E}\{\mathbf{g}[r]\mathbf{g}[r']\} + o\left( \frac{1}{K} \right) \; . \tag{60}$$

Thus, given the bounds (38) on $|\mathbb{E}\{\mathbf{g}[r]\mathbf{g}[r']\}|$ and (48) on the second derivative of $f_m^{(\ell)}$, we have:

$$|\epsilon_2[\ell]| \leq C_\mathbf{z} \frac{M^3(M+1)^2}{4} \left\| \left. \frac{\partial^2 f^{(\ell)}}{\partial \mathbf{g}[r] \partial \mathbf{g}[r']} \right|_{\mathbf{g}=0} \right\|_{\max} \frac{1}{K} \left( C_\mathbf{z}^2 \sigma^2 + 2\sigma^4 \right) + o\left( \frac{1}{K} \right)$$

$$\leq \frac{1}{K} \left( c_1^{(2)} + \frac{c_2^{(2)}}{\sigma^2} \right) + o\left( \frac{1}{K} \right) \; , \tag{61}$$

where

$$c_1^{(2)} \triangleq \frac{M^3(M+1)^2}{2} C_\mathbf{z} d_{2,\mathbf{z},M,\ell} \; ,$$

$$c_2^{(2)} \triangleq \frac{M^3(M+1)^2}{4} C_\mathbf{z}^3 d_{2,\mathbf{z},M,\ell} \; .$$

Let us now determine an upper bound on $|\epsilon_3[\ell]|$. A second-order Taylor expansion of $\mathbf{h}^{(\ell)}$ gives

$$\mathbb{E}\{\mathbf{h}^{(\ell)} \mathbf{w}_K\} = \sum_{m=0}^{M-1} \sum_{r=0}^{M(M+1)-1} \left. \frac{\partial f_m^{(\ell)}}{\partial \mathbf{g}[r]} \right|_{\mathbf{g}=0} \mathbb{E}\{\mathbf{g}[r]\mathbf{w}_K[m]\} + o\left( \frac{1}{K} \right) \tag{62}$$

Indeed, the third-order moments $\mathbb{E}\{\mathbf{g}[r]\mathbf{g}[r']\mathbf{w}_K[m]\}$, behave as $o\left( \frac{1}{K} \right)$. Besides,

$$\mathbb{E}\{\mathbf{g}[r]\mathbf{w}_K[m]\} = \sigma \mathbb{E}\{\mathbf{g}_1[r]\mathbf{w}_K[m]\} + \sigma^2 \mathbb{E}\{\mathbf{g}_2[r]\mathbf{w}_K[m]\} \; . \tag{63}$$

Then,

$$|\mathbb{E}\{\mathbf{g}_1[r]\mathbf{w}_K[m]\}| = \frac{1}{K} \left| \sum_{k=0}^{K-1} \mathbf{z}_k[m_r + a_r] \mathbb{E}\{\mathbf{w}_k[m_r']\mathbf{w}_K[m]\} + \mathbf{z}_k[m_r'] \mathbb{E}\{\mathbf{w}_k[m_r + a_r]\mathbf{w}_K[m]\} \right|$$

$$\leq \frac{2}{K} \left( \sum_{j=1}^{J} \Omega_j \right) = \frac{C_\mathbf{z}}{K} \; . \tag{64}$$

$$\mathbb{E}\{\mathbf{g}_2[r]\mathbf{w}_K[m]\} = \frac{1}{K}\sum_{k,=0}^{K-1}\mathbb{E}\left\{\mathbf{w}_k[m_r+a_r]\mathbf{w}_k[m_r']\mathbf{w}_K[m]\right\} - \delta_{m_r+a_r,m_r'}\mathbb{E}\left\{\mathbf{w}_K[m]\right\}$$

$$= 0 . \tag{65}$$

Thus, combining results (64) and (65) into expression (63) gives the following bound:

$$|\mathbb{E}\{\mathbf{g}[r]\mathbf{w}_K[m]\}| \leq C_{\mathbf{z}}\frac{\sigma}{K} . \tag{66}$$

Thus, given the bound (47) on the first derivative of $f_m^{(\ell)}$, and from (62) we have:

$$|\epsilon_3[\ell]| \leq \sigma M^2(M+1)\frac{d_{1,\mathbf{z},M,\ell}}{\sigma^2}C_{\mathbf{z}}\frac{\sigma}{K} \triangleq \frac{c^{(3)}}{K} , \tag{67}$$

where $c^{(3)} \triangleq M^2(M+1)d_{1,\mathbf{z},M,\ell}C_{\mathbf{z}}$.

Thus, combining bounds (59) on $\epsilon_1$, (61) on $\epsilon_2$ and (67) on $\epsilon_3$ gives the following bound of the bias:

$$|\boldsymbol{\mu}[n]| \leq c^{(1)}\sigma^2 + \frac{1}{K}\left(\frac{c_2^{(2)}}{\sigma^2} + c_1^{(2)} + c^{(3)}\right) + o\left(\frac{1}{K}\right) . \tag{68}$$

*C. Expression of the Covariance $\gamma$.*

Outside the measurement interval $I$, denote by $\ell$ the index such that $n = N - 1 + \ell$. Then, given that $\mathbf{h}^{(\ell)} = \boldsymbol{\alpha}^{(\ell)} - \boldsymbol{\alpha}_0^{(\ell)}$, we have

$$\gamma[n,n] = \left(\boldsymbol{\alpha}_0^{(\ell)}\mathbf{z}_K\right)^2 + 2\left(\boldsymbol{\alpha}_0^{(\ell)}\mathbf{z}_K\right)\mathbb{E}\left\{\mathbf{h}^{(\ell)}\right\}\mathbf{z}_K + \mathbb{E}\left\{\left(\mathbf{h}^{(\ell)}\mathbf{z}_K\right)^2\right\} + 2\sigma\boldsymbol{\alpha}_0^{(\ell)}\mathbb{E}\{\mathbf{w}_K\mathbf{h}^{(\ell)}\}\mathbf{z}_K + 2\sigma\boldsymbol{\alpha}_0^{(\ell)}\mathbf{z}_K\mathbb{E}\{\mathbf{h}^{(\ell)}\mathbf{w}_K\}$$

$$+ 2\sigma\mathbb{E}\{\mathbf{h}^{(\ell)}\mathbf{w}_K\mathbf{h}^{(\ell)}\}\mathbf{z}_K + \sigma^2\left\|\boldsymbol{\alpha}_0^{(\ell)}\right\|^2 + \sigma^2\mathbb{E}\left\{\left(\mathbf{h}^{(\ell)}\mathbf{w}_K\right)^2\right\} + 2\sigma^2\boldsymbol{\alpha}_0^{(\ell)}\mathbb{E}\{\mathbf{w}_K\mathbf{h}^{(\ell)}\mathbf{w}_K\} - \mathbf{z}[n]^2 - 2\mathbf{z}[n]\boldsymbol{\mu}[n] - \boldsymbol{\mu}[n]^2$$

$$= \sigma^2\left\|\boldsymbol{\alpha}_0^{(\ell)}\right\|^2 - \left(\mathbf{z}[n] - \boldsymbol{\alpha}_0^{(\ell)}\mathbf{z}_K + \mathbf{z}[n]\right)^2 + \mathbb{E}\left\{\left(\mathbf{h}^{(\ell)}\mathbf{z}_K\right)^2\right\} + 2\sigma\boldsymbol{\alpha}_0^{(\ell)}\mathbb{E}\{\mathbf{w}_K\mathbf{h}^{(\ell)}\}\mathbf{z}_K + 2\sigma\mathbb{E}\{\mathbf{h}^{(\ell)}\mathbf{w}_K\mathbf{h}^{(\ell)}\}\mathbf{z}_K$$

$$+ \sigma^2\mathbb{E}\left\{\left(\mathbf{h}^{(\ell)}\mathbf{w}_K\right)^2\right\} + 2\sigma^2\boldsymbol{\alpha}_0^{(\ell)}\mathbb{E}\{\mathbf{w}_K\mathbf{h}^{(\ell)}\mathbf{w}_K\}$$

$$\triangleq \eta_1[n] + \eta_2[n] + \eta_3[n] + \eta_4[n] + \eta_5[n] + \eta_6[n] + \eta_7[n] ,$$

where

$$\eta_1[n] = \sigma^2\left\|\boldsymbol{\alpha}_0^{(\ell)}\right\|^2 , \qquad \eta_2[n] = -\left(\mathbf{z}[n] - \boldsymbol{\alpha}_0^{(\ell)}\mathbf{z}_K + \mathbf{z}[n]\right)^2 , \qquad \eta_3[n] = \mathbb{E}\left\{\left(\mathbf{h}^{(\ell)}\mathbf{z}_K\right)^2\right\} ,$$

$$\eta_4[n] = 2\sigma\boldsymbol{\alpha}_0^{(\ell)}\mathbb{E}\{\mathbf{w}_K\mathbf{h}^{(\ell)}\}\mathbf{z}_K , \qquad \eta_5[n] = \sigma\mathbb{E}\{\mathbf{h}^{(\ell)}\mathbf{w}_K\mathbf{h}^{(\ell)}\}\mathbf{z}_K , \qquad \eta_6[n] = \sigma^2\mathbb{E}\left\{\left(\mathbf{h}^{(\ell)}\mathbf{w}_K\right)^2\right\} ,$$

$$\eta_7[n] = 2\sigma^2\boldsymbol{\alpha}_0^{(\ell)}\mathbb{E}\{\mathbf{w}_K\mathbf{h}^{(\ell)}\mathbf{w}_K\} .$$

Let us now determine an upper bound on each of these terms.

First, since $\|\boldsymbol{\alpha}_0^{(\ell)}\|^2 \leq M\|\mathbf{A}_0^\ell\|_{\max}^2$, we have:

$$\eta_1[n] \leq \sigma^2 M\left\|\mathbf{A}_0^\ell\right\|_{\max}^2$$

$$\leq \sigma^2 M^\ell\|\mathbf{A}_0\|_{\max}^{2\ell}$$

$$\leq \sigma^2 M^\ell \max\left(1, \left(\frac{2J}{M}\right)^{2\ell}\right) . \tag{69}$$

Second, by definition of $\epsilon_2$ and $\epsilon_3$ (see expressions (55) and (56)), $\eta_2$ takes the following form:

$$\eta_2[n] = (\epsilon_2[n] + \epsilon_3[n])^2 = o\left(\frac{1}{K}\right) . \tag{70}$$

Third, second-order Taylor expansions of $\mathbf{h}^{(\ell)}$ give:

$$\eta_3[n] \leq \frac{C_{\mathbf{z}}^2}{4} \sum_{m,m'=0}^{M-1} \left| \mathbb{E}\left\{\mathbf{h}^{(\ell)}[m]\mathbf{h}^{(\ell)}[m']\right\} \right|$$

$$\leq \frac{C_{\mathbf{z}}^2}{4} \sum_{r,r'=0}^{M(M+1)-1} \left| \frac{\partial f_m^{(\ell)}}{\partial \mathbf{g}[r]}\Bigg|_{\mathbf{g}=\mathbf{0}} \frac{\partial f_{m'}^{(\ell)}}{\partial \mathbf{g}[r']}\Bigg|_{\mathbf{g}=\mathbf{0}} \mathbb{E}\{\mathbf{g}[r]\mathbf{g}[r']\} \right| + o\left(\frac{1}{K}\right)$$

$$\leq \frac{C_{\mathbf{z}}^2 M^2(M+1)^2}{4} \frac{d_{1,\mathbf{z},M,\ell}^2}{\sigma^4} \frac{1}{K}\left(C_{\mathbf{z}}^2\sigma^2 + 2\sigma^4\right) + o\left(\frac{1}{K}\right)$$

$$\leq \frac{C_{\mathbf{z}}^2 M^2(M+1)^2 d_{1,\mathbf{z},M,\ell}^2}{4K}\left(\frac{C_{\mathbf{z}}^2}{\sigma^2} + 2\right) + o\left(\frac{1}{K}\right) \; . \tag{71}$$

Fourth,

$$|\eta_4[n]| \leq \sigma C_{\mathbf{z}} \left\|\mathbf{A}_0^\ell\right\|_{\max} \sum_{m,m'=0}^{M-1} \left|\mathbb{E}\{\mathbf{h}^{(\ell)}[m]\mathbf{w}_K[m']\}\right|$$

But,

$$\left|\mathbb{E}\{\mathbf{h}^{(\ell)}[m]\mathbf{w}_K[m']\}\right| \leq \sum_{r=0}^{M(M+1)-1} \left| \frac{\partial f_m^{(\ell)}}{\partial \mathbf{g}[r]}\Bigg|_{\mathbf{g}=\mathbf{0}} \right| \left|\mathbb{E}\{\mathbf{g}[r]\mathbf{w}_K[m']\}\right| + o\left(\frac{1}{K}\right)$$

$$\leq M(M+1)\frac{d_{1,\mathbf{z},M,\ell}}{\sigma^2}C_{\mathbf{z}}\frac{\sigma}{K} + o\left(\frac{1}{K}\right) \tag{72}$$

Thus,

$$|\eta_4[n]| \leq M^\ell(M+1)C_{\mathbf{z}}^2 d_{1,\mathbf{z},M,\ell}\left(\max\left(1,\frac{2J}{M}\right)\right)^\ell \frac{1}{K} + o\left(\frac{1}{K}\right) \; . \tag{73}$$

Fifth and sixth, second-order Taylor expansions of $\mathbf{h}^{(\ell)}$ give

$$\eta_5[n] = o\left(\frac{1}{K}\right) \; , \tag{74}$$

$$\eta_6[n] = o\left(\frac{1}{K}\right) \; . \tag{75}$$

Seventh,

$$|\eta_7[n]| \leq 2\sigma^2 \left\|\mathbf{A}_0^\ell\right\|_{\max} \sum_{m,m'=0}^{M-1} \left|\mathbb{E}\{\mathbf{h}^{(\ell)}[m]\mathbf{w}_K[m]\mathbf{w}_K[m']\}\right|$$

$$\leq 2\sigma^2 \left\|\mathbf{A}_0^\ell\right\|_{\max} \sum_{m,m'=0}^{M-1} \sum_{r=0}^{M(M+1)-1} \left| \frac{\partial f_m^{(\ell)}}{\partial \mathbf{g}[r]}\Bigg|_{\mathbf{g}=\mathbf{0}} \right| \left|\mathbb{E}\{\mathbf{g}^{(\ell)}[r]\mathbf{w}_K[m]\mathbf{w}_K[m']\}\right| + o\left(\frac{1}{K}\right) \; .$$

But,

$$\mathbb{E}\{\mathbf{g}^{(\ell)}[r]\mathbf{w}_K[m]\mathbf{w}_K[m']\} = \sigma\mathbb{E}\{\mathbf{g}_1^{(\ell)}[r]\mathbf{w}_K[m]\mathbf{w}_K[m']\} + \sigma^2\mathbb{E}\{\mathbf{g}_2^{(\ell)}[r]\mathbf{w}_K[m]\mathbf{w}_K[m']\} \; .$$

As before, since $\mathbb{E}\{\mathbf{g}_1^{(\ell)}[r]\mathbf{w}_K[m]\mathbf{w}_K[m']\}$ is a third-order moment of a multivariate zero-mean Gaussian vector, it vanishes. And,

$$\left|\mathbb{E}\{\mathbf{g}_2^{(\ell)}[r]\mathbf{w}_K[m]\mathbf{w}_K[m']\}\right| = \frac{1}{K}\left|\sum_{k=0}^{K-1} \mathbb{E}\left\{\mathbf{w}_k[m_r+a_r]\mathbf{w}_k[m'_r]\mathbf{w}_K[m]\mathbf{w}_K[m']\right\} - \delta_{m_r+a_r,m'_r}\mathbb{E}\left\{\mathbf{w}_K[m]\mathbf{w}_K[m']\right\}\right|$$

$$= \frac{1}{K}\left|\sum_{k=0}^{K-1} \delta_{k+m_r+a_r,K+m}\delta_{k+m'_r,K+m'} + \delta_{k+m_r+a_r,K+m'}\delta_{k+m'_r,K+m}\right|$$

$$\leq \frac{2}{K} \; .$$

Thus,

$$|\eta_7[n]| \leq 2\sigma^2 \left\|\mathbf{A}_0^\ell\right\|_{\max} M^3(M+1)\frac{d_{1,\mathbf{z},M,\ell}}{\sigma^2}\frac{2\sigma^2}{K} + o\left(\frac{1}{K}\right)$$

$$\leq 4M^{\ell+2}(M+1)d_{1,\mathbf{z},M,\ell}\left(\max\left(1,\frac{2J}{M}\right)\right)^\ell\frac{\sigma^2}{K} + o\left(\frac{1}{K}\right). \tag{76}$$

To conclude, we combine the expressions (69), (70), (71), (73), (74), (75), and (76) to determine an upper bound on the variance $\gamma[n,n]$. It follows:

$$\gamma[n,n] \leq c_0^{(n)}\sigma^2 + \frac{1}{K}\left(\frac{c_1^{(n)}}{\sigma^2} + c_2^{(n)} + c_3^{(n)}\sigma^2\right) + o\left(\frac{1}{K}\right),$$

where

$$c_0^{(n)} = M^\ell \max\left(1,\left(\frac{2J}{M}\right)^{2\ell}\right)$$

$$c_1^{(n)} = \frac{1}{4}C_\mathbf{z}^4 M^2(M+1)^2 d_{1,\mathbf{z},M,\ell}^2$$

$$c_2^{(n)} = \frac{1}{2}C_\mathbf{z}^2 M^2(M+1)^2 d_{1,\mathbf{z},M,\ell}^2 + M^\ell(M+1)C_\mathbf{z}^2 d_{1,\mathbf{z},M,\ell}\left(\max\left(1,\frac{2J}{M}\right)\right)^\ell$$

$$c_3^{(n)} = 4M^{\ell+2}(M+1)d_{1,\mathbf{z},M,\ell}\left(\max\left(1,\frac{2J}{M}\right)\right)^\ell.$$

*a) If $n' \geq N$:* When $n \geq N$ and $n' \geq N$, applying the Cauchy-Schwarz inequality, we obtain the following bound on the covariance $\gamma[n,n]$:

$$|\gamma[n,n']| \leq \sqrt{\gamma[n,n]\gamma[n',n']}$$

$$\leq \sqrt{c_0^{(n)}c_0^{(n')}}\sigma^2\sqrt{1+\frac{1}{K\sigma^2}\left(\frac{1}{c_0^{(n)}}\left(\frac{c_1^{(n)}}{\sigma^2}+c_2^{(n)}+c_3^{(n)}\sigma^2\right)+\frac{1}{c_0^{(n')}}\left(\frac{c_1^{(n')}}{\sigma^2}+c_2^{(n')}+c_3^{(n')}\sigma^2\right)\right)}+o\left(\frac{1}{K}\right).$$

A first-order Taylor expansion of this bound as $K \to \infty$ gives

$$\leq \sqrt{c_0^{(n)}c_0^{(n')}}\sigma^2 + \frac{1}{2K}\left(\sqrt{\frac{c_0^{(n')}}{c_0^{(n)}}}\left(\frac{c_1^{(n)}}{\sigma^2}+c_2^{(n)}+c_3^{(n)}\sigma^2\right)+\sqrt{\frac{c_0^{(n)}}{c_0^{(n')}}}\left(\frac{c_1^{(n')}}{\sigma^2}+c_2^{(n')}+c_3^{(n')}\sigma^2\right)\right)+o\left(\frac{1}{K}\right)$$

$$\leq c_0^{(n,n')}\sigma^2 + \frac{1}{K}\left(\frac{c_1^{(n,n')}}{\sigma^2}+c_2^{(n,n')}+c_3^{(n,n')}\sigma^2\right)+o\left(\frac{1}{K}\right),$$

where

$$c_0^{(n,n')} = \sqrt{c_0^{(n)}c_0^{(n')}},$$

$$c_p^{(n,n')} = \frac{1}{2}\left(c_p^{(n)}\sqrt{\frac{c_0^{(n')}}{c_0^{(n)}}}+c_p^{(n')}\sqrt{\frac{c_0^{(n)}}{c_0^{(n')}}}\right), \quad \forall p \in \{1,2,3\}.$$

*b) If $n' \in I$:* When $n \geq N$ and $n' \in I$, equation (17) shows us that:

$$\gamma[n,n'] = \sigma\mathbb{E}\{\mathbf{w}[n']\mathbf{h}^{(\ell)}\}\mathbf{z}_K + \sigma^2\mathbb{E}\{\mathbf{w}[n']\boldsymbol{\alpha}_0^{(\ell)}\mathbf{w}_K\} + \sigma^2\mathbb{E}\{\mathbf{w}[n']\mathbf{h}^{(\ell)}\mathbf{w}_K\}$$

$$\stackrel{\Delta}{=} \beta_1[n,n'] + \beta_2[n,n'] + \beta_3[n,n'],$$

where

$$\beta_1[n,n'] = \sigma\mathbb{E}\{\mathbf{w}[n']\mathbf{h}^{(\ell)}\}\mathbf{z}_K,$$

$$\beta_2[n,n'] = \sigma^2\mathbb{E}\{\mathbf{w}[n']\boldsymbol{\alpha}_0^{(\ell)}\mathbf{w}_K\},$$

$$\beta_3[n,n'] = \sigma^2\mathbb{E}\{\mathbf{w}[n']\mathbf{h}^{(\ell)}\mathbf{w}_K\}.$$

Besides, thanks to the bound (72) we have:

$$\left|\beta_1[n,n']\right| \leq \sigma \frac{C_{\mathbf{z}}}{2} \sum_{m=0}^{M-1} \left|\mathbb{E}\{\mathbf{w}[n']\mathbf{h}^{(\ell)}[m]\}\right|$$

$$\leq \sigma \frac{C_{\mathbf{z}}}{2} M^2(M+1) \frac{d_{1,\mathbf{z},M,\ell}}{\sigma^2} C_{\mathbf{z}} \frac{\sigma}{K} + o\left(\frac{1}{K}\right)$$

$$\leq M^2(M+1) \frac{C_{\mathbf{z}}^2 d_{1,\mathbf{z},M,\ell}}{2} \frac{1}{K} + o\left(\frac{1}{K}\right) . \tag{77}$$

Besides,

$$\left|\beta_2[n,n']\right| \leq \sigma^2 \left\|\mathbf{A}_0^\ell\right\|_{\max} \sum_{m=0}^{M-1} \left|\mathbb{E}\{\mathbf{w}[n']\mathbf{w}_K[m]\}\right|$$

$$\leq \sigma^2 \left\|\mathbf{A}_0^\ell\right\|_{\max}$$

$$\leq \sigma^2 M^{\ell-1} \left(\max\left(1, \frac{2J}{M}\right)\right)^\ell . \tag{78}$$

Besides, identically to the bound (76) on $\eta_7$, we obtain

$$\left|\beta_3[n,n']\right| \leq \sigma^2 \sum_{m=0}^{M-1} \left|\mathbb{E}\{\mathbf{w}[n']\mathbf{h}^{(\ell)}[m]\mathbf{w}_K[m]\}\right|$$

$$\leq \sigma^2 M^3(M+1) \frac{d_{1,\mathbf{z},M,\ell}}{\sigma^2} \frac{4\sigma^2}{K} + o\left(\frac{1}{K}\right)$$

$$\leq M^3(M+1) d_{1,\mathbf{z},M,\ell} \frac{4\sigma^2}{K} + o\left(\frac{1}{K}\right) . \tag{79}$$

Finally, we combine expressions (77), (78), and (79) to determine an upper bound on the variance $\gamma[n,n']$. It follows:

$$\left|\gamma[n,n']\right| \leq b_0^{(n,n')}\sigma^2 + \frac{1}{K}\left(b_1^{(n,n')} + b_2^{(n)}\sigma^2\right) + o\left(\frac{1}{K}\right) ,$$

where

$$b_0^{(n,n')} = M^{\ell-1}\left(\max\left(1, \frac{2J}{M}\right)\right)^\ell$$

$$b_1^{(n,n')} = M^2(M+1)\frac{C_{\mathbf{z}}^2 d_{1,\mathbf{z},M,\ell}}{2}$$

$$b_2^{(n,n')} = 4M^3(M+1)d_{1,\mathbf{z},M,\ell} .$$

## II. APPLICATION TO AN ELECTROCARDIOGRAM

We provide here an additional implementation of BoundEffRed, applied to an electrocardiogram (ECG) dataset. The dataset is constructed from a 500-second-long ECG, sampled at $f_{\mathbf{s}} = 200$ Hz, cut into 10 segments of 50 seconds each. Fig. 10 depicts the right boundary of one of these subsignals, together with the 6-second extensions estimated by SigExt (first panel), or EDMD (second panel), GPR (third panel), or TBATS (fourth panel). These extensions are superimposed to the ground-truth extension, plotted in red. The sharp and spiky ECG patterns make the AHM model too simplistic to describe this type of signal. Consequently, the forecast produced by SigExt is moderately satisfactory. Note that TBATS is the only one that seems to accurately capture the locations of QRS complexes after 37 seconds. We will explore this long-term prediction capability in the future work.

Table IV contains the median performance index $D$ of the boundary-free TF representations, over the $N$ subsignals evaluated, according to the extension method. As a result of the fair quality of the forecasts, the reduction of boundary effects is less significant than for PPG signal. Nevertheles, the results show that BoundEffRed has the same efficiency when the SigExt extension, the EDMD extension or the GPR extension is chosen. Indeed, t-tests performed under the null hypothesis that the mean are equals, with a 5% significance level, show no statistical significant difference between SigExt and EDMD or GPR, regardless of the representation considered. This justifies the choice of SigExt for real-time implementation.
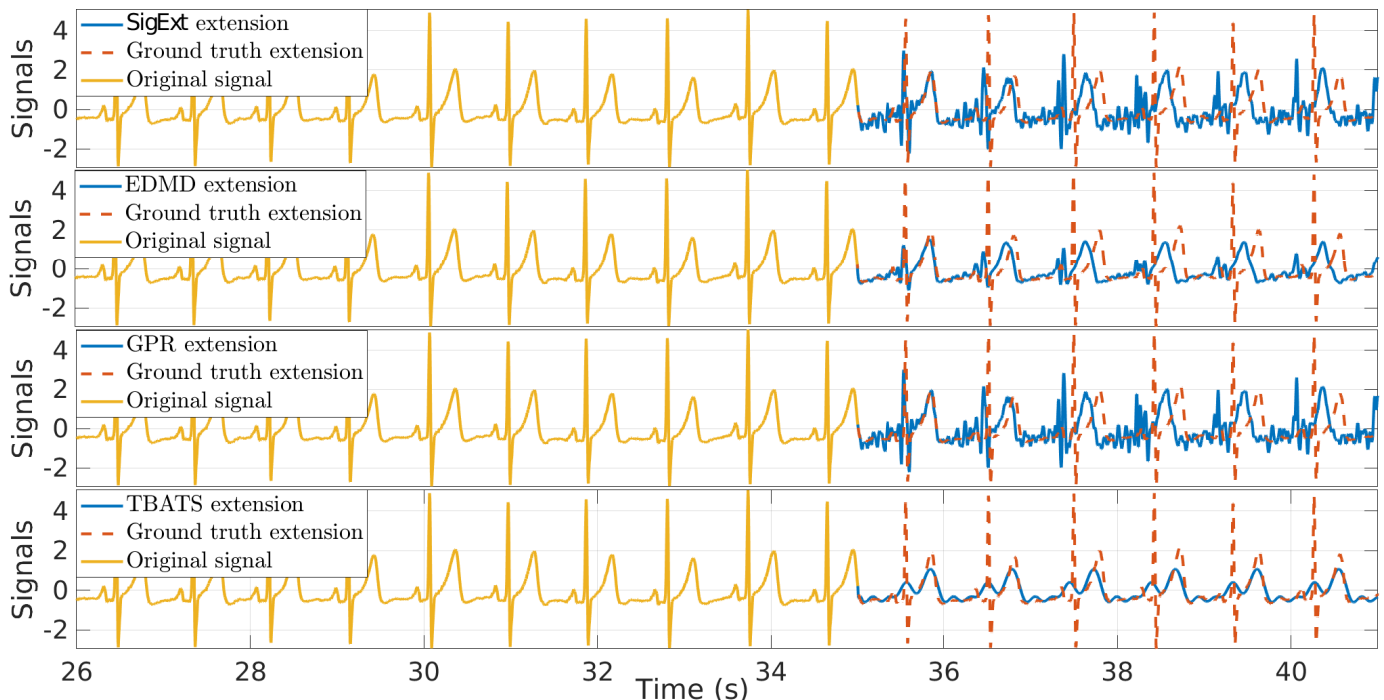
Fig. 10. Extended ECG (blue) obtained by the SigExt forecasting (first panel), the EDMD forecasting (second panel), the GPR forecasting (third panel), and the TBATS forecasting (fourth panel), superimposed with the ground truth signal (dashed red).

TABLE IV
ECG: PERFORMANCE OF THE BOUNDARY-FREE TF REPRESENTATIONS ACCORDING TO THE EXTENSION METHOD

| Extension method | Median performance index $D$ | | | |
|---|---|---|---|---|
| | STFT | SST | ConceFT | RS |
| SigExt | 0.584 | 0.630 | 0.462 | 0.642 |
| Symmetric | 1.199 | 1.354 | 1.427 | 0.943 |
| EDMD | 0.538 | 0.558 | 0.496 | 0.714 |
| GPR | 0.639 | 0.588 | 0.485 | 0.616 |

## III. APPLICATION TO A MULTICOMPONENT CARDIAC SIGNAL

We consider a cardiac signal, namely a photoplethysmogram (PPG) recording, sampled at 300 Hz. In addition to the cardiac cycle measurement, this signal contains a slow varying component, which is the respiratory component, known as respiration induced intensity variation (RIIV) [2]. The top of Fig. 11 displays an excerpt of the signal along with a 3-second-long extension obtained by the SigExt forecasting. The lower part of Fig. 11 shows the respiratory signal recording the concentration of $CO_2$. This signal was recorded simultaneously with the PPG signal, and visually highlights the low-frequency respiratory component contained in the PPG signal. Indeed, the intervals where the concentration of $CO_2$ drops—highlighted by the bluish areas—coincide with the decreases in the PPG signal. Note also that the forecasting breaks the waveform of the oscillations because of its inability to forecast the high-frequency harmonics contained in the signal. Nevertheless, as long as the low-frequency harmonics of the components contained in the signal are preserved, the forecasting is sufficient to reduce boundary effects in the low-frequency TF domain.

The ordinary and boundary-free SSTs of this signal are displayed in Fig. 12. These TF analyses brings out both components—the fundamental frequency of the cardiac component, the most energetic one, indicated by the blue arrows, is located around 1.3 Hz, while the respiratory component, less visible, indicated by the green arrows, is located around 0.2 Hz and its multiples. Clearly, near the boundaries, BoundEffRed helps improve the quality of the TF representation. Besides, the performance index of this representation takes the value $D = 0.491$. This means that BoundEffRed has reduced the right-side boundary effects by about 51% with respect to the ordinary SST. This shows the ability of our algorithm to work on signals containing several nonstationary components.

## REFERENCES

[1] L. Isserlis, "On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables," *Biometrika*, vol. 12, no. 1-2, pp. 134–139, 11 1918. [Online]. Available: https://doi.org/10.1093/biomet/12.1-2.134

[2] L. M. Nilsson, "Respiration signals from photoplethysmography," *Anesthesia and Analgesia*, vol. 117, no. 4, pp. 859–865, 2013.
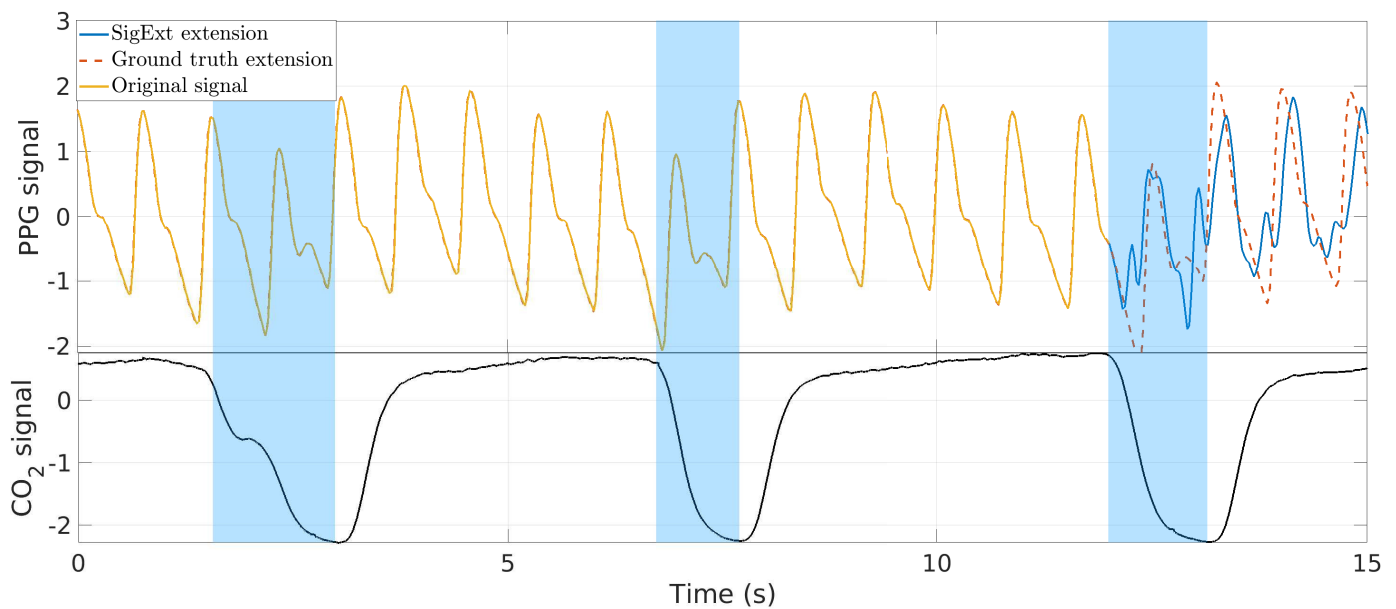
Fig. 11. PPG signal (top, yellow), and the associated SigExt extension (blue). The simultaneously recorded concentration of $CO_2$ is below. The bluish areas show the synchrony between the drops in $CO_2$ concentration and the decreases in the PPG signal.
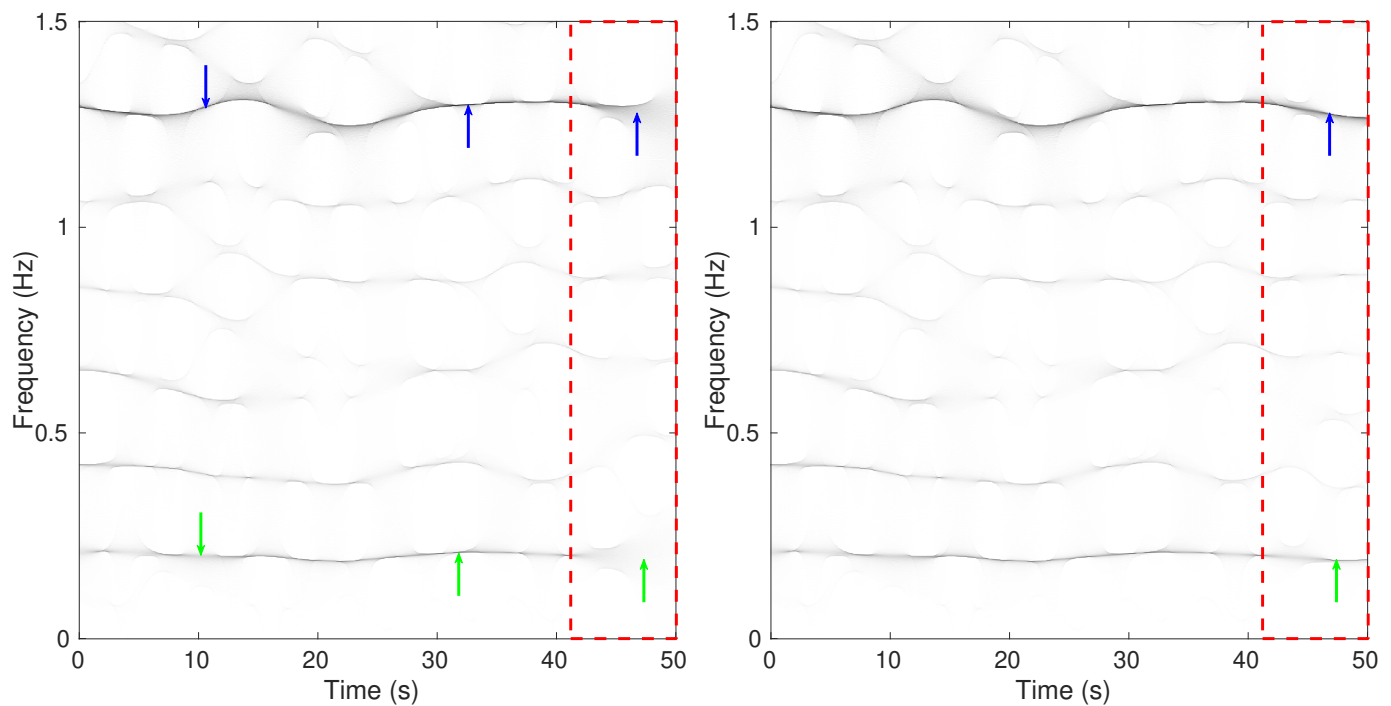


Fig. 12. Ordinary SST (left) and boundary-free SST (right) of the PPG signal. The window length for the SSTs is 18 seconds. The green arrows indicate the instantaneous frequency of the respiratory component. The blue arrows, on the other hand, indicate the instantaneous frequency of the blood pressure component. The red dashed boxes highlight the areas where boundary effects may appear.