Optimality property of the Gaussian window spectrogram

Citation for published version (APA):

DOI:
10.1109/78.80783

Document status and date:
Published: 01/01/1991

Document Version:
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher’s website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:
openaccess@tue.nl

providing details and we will investigate your claim.
TABLE I
SUMMARY OF ADAPTATION

<table>
<thead>
<tr>
<th>Adaptation</th>
<th>Delay</th>
<th>Gain</th>
<th>Average Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20</td>
<td>0.498163</td>
<td>0.677243</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>0.447455</td>
<td>0.588629</td>
</tr>
<tr>
<td>3</td>
<td>18</td>
<td>-0.402277</td>
<td>0.513097</td>
</tr>
<tr>
<td>4</td>
<td>13</td>
<td>-0.367075</td>
<td>0.450902</td>
</tr>
<tr>
<td>5</td>
<td>22</td>
<td>-0.411588</td>
<td>0.393350</td>
</tr>
<tr>
<td>6</td>
<td>17</td>
<td>-0.390106</td>
<td>0.328782</td>
</tr>
<tr>
<td>7</td>
<td>23</td>
<td>-0.329475</td>
<td>0.258251</td>
</tr>
<tr>
<td>8</td>
<td>25</td>
<td>0.318380</td>
<td>0.227739</td>
</tr>
<tr>
<td>9</td>
<td>12</td>
<td>-0.228103</td>
<td>0.200542</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>0.246870</td>
<td>0.166498</td>
</tr>
</tbody>
</table>

The author is with Philips Research Laboratories, The Netherlands.

The Netherlands.

IEEE Log Number 9040410.

Fig. 5. Magnitude spectra with plant noise 1:1 in adaptive filter (30 terms) and adaptive delay filter (10 terms).

Fig. 6. Magnitude spectra of adaptive filter (30 terms) and adaptive delay filter (10 terms).

Fig. 7. Magnitude spectra with plant noise 1:1 in adaptive filter (30 terms) and adaptive delay filter (10 terms).

number of terms, or until a specified tolerance mean-squared error value is achieved. With a standard adaptive filter, the model order must be selected before the adaptation begins.

Another advantage with the adaptive delay filter is that we obtain the optimal FIR model for a specified number of filter components. For example, suppose we want the optimal model for a system using only 10 components. The adaptive delay filter with 10 elements directly computes this model. If we select the ten largest components from a Wiener solution or a standard adaptive filter model, we do not have an optimal solution for 10 components.

An additional advantage of the adaptive delay filter is that it requires less data than standard adaptive filter solutions. Since the adaptive delay filter finds each delay/gain component sequentially, the same set of data can be used for the adaptation of each delay/gain component. (The "desired signal" for each component becomes the "error signal" from the previous adaptation.) The example adaptation of the adaptive delay filter with 10 terms in the previous section was done with only 380 data points.

In summary, if a sparse or reduced model is needed for a known system or an unknown system, the adaptive delay filter provides an efficient technique for finding the optimal model using only a small window of data.

REFERENCES


Optimality Property of the Gaussian Window Spectrogram

A. J. E. M. Janssen

Abstract—It is shown that for any signal \( x(t) \) the minimum of
\[
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[ (t - t_0)^2 + (f - f_0)^2 \right] S_x^c(t,f) \, dt \, df
\]

Manuscript received July 18, 1988; revised May 16, 1989.

The author is with Philips Research Laboratories, 5600 JA Eindhoven, The Netherlands.

IEEE Log Number 9040410.
over all normalized time-windows \( w(t) \) is achieved by the Gaussian window \( w(t) = 2^{1/4} \exp(-\pi t^2) \). Here, \( (t_0, f_0) \) is the center of gravity of the signal \( x(t) \), \( S_{t_0}^{(w)}(t, f) \) is the spectrogram of \( x(t) \) due to the window \( w(t) \), and the double integral is a measure of the spread of \( S_{t_0}^{(w)}(t, f) \) around \( (t_0, f_0) \) in the time-frequency plane.

I. INTRODUCTION

When \( x(t) \) is a square integrable signal and \( w(t) \) is a square integrable window, the spectrogram of \( x(t) \) due to \( w(t) \) is defined by

\[
S_{t_0}^{(w)}(t, f) = \left| \int_{-\infty}^{\infty} x(r) w(r-t) e^{-2 \pi ifr} dr \right|^2
\]  (1.1)

or, equivalently, by

\[
S_{t_0}^{(w)}(t, f) = \left| \int_{-\infty}^{\infty} X(g) W(f-g) e^{2 \pi ifg} dg \right|^2
\]  (1.2)

with \( X(f) \) and \( W(f) \) being the Fourier transforms of \( x(t) \) and \( w(t) \), respectively. We use here the definition

\[ y(f) = y(t) e^{-2 \pi ift} \]  (1.3)

for the Fourier transform \( Y(f) \) of \( y(t) \). In case \( x(t) \) is a signal whose energy is well concentrated around a point \( (t_0, f_0) \) in the time-frequency plane, one would like \( S_{t_0}^{(w)}(t, f) \) to be well-concentrated around \( (t_0, f_0) \) as well. As a measure for this, one could take

\[
\Sigma_{t_0}^{(w)}(t_0, f_0) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[ (t-t_0)^2 + (f-f_0)^2 \right] S_{t_0}^{(w)}(t, f) \, dt \, df
\]  (1.4)

which is a quadratic time-frequency moment around \( (t_0, f_0) \). In this correspondence we shall show the following. Assume that

\[ \int_{-\infty}^{\infty} |x(t)|^2 \, dt = \int_{-\infty}^{\infty} |X(f)|^2 \, df = 1 \]  (1.5)

and let

\[ t_0 = \int_{-\infty}^{\infty} t |x(t)|^2 \, dt, \quad f_0 = \int_{-\infty}^{\infty} f |X(f)|^2 \, df \]  (1.6)

so that \( (t_0, f_0) \) is the center of gravity of \( x(t) \). Then \( \Sigma_{t_0}^{(w)}(t_0, f_0) \) is, among all \( w(t) \) with

\[ \int_{-\infty}^{\infty} |w(t)|^2 \, dt = \int_{-\infty}^{\infty} |W(f)|^2 \, df = 1 \]  (1.7)

uniquely minimized by the Gaussian

\[ w(t) = 2^{1/4} \exp(-\pi t^2) \]  (1.8)

with \( c \) a complex constant with \( |c| = 1 \).

This result is somewhat remarkable since the optimal \( w(t) \) is completely independent of \( x(t) \). We prove our result by relating \( \Sigma_{t_0}^{(w)}(t_0, f_0) \) to a different measure of concentration in the time-frequency plane, viz., to

\[ \sigma_{t_0}^2(r, g) = \int_{-\infty}^{\infty} (r - t_0)^2 |Y(t)|^2 \, dt + \int_{-\infty}^{\infty} (g - f_0)^2 |Y(f)|^2 \, df \]  (1.9)

where \( y \) is any signal and \( (r, g) \) is any point in the time-frequency plane. We shall show that

\[
\Sigma_{t_0}^{(w)}(t_0, f_0) = \sigma_{t_0}^2(t_0, f_0) + \sigma_{0}^2(0, 0) \]  (1.10)

so that minimization of \( \Sigma_{t_0}^{(w)}(t_0, f_0) \) reduces to minimization of \( \sigma_{t_0}^2(0, 0) \), and that \( \sigma_{0}^2(0, 0) \) is uniquely minimal for the Gaussian \( w(t) \) of (1.8).

II. DERIVATIONS

We shall prove formula (1.10) under the assumptions (1.5) and (1.7). We have from (1.4)

\[
\Sigma_{t_0}^{(w)}(t_0, f_0) = \Sigma_{t_0}^{(w)}(t_0) + \Sigma_{t_0}^{(w)}(f_0) \]  (2.1)

where

\[
\Sigma_{t_0}^{(w)}(t_0) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (r - t_0)^2 S_{t_0}^{(w)}(r, f) \, dr \, df \]  (2.2)

\[ \Sigma_{t_0}^{(w)}(f_0) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (f - f_0)^2 S_{t_0}^{(w)}(r, f) \, dr \, df \]  (2.3)

Consider \( \Sigma_{t_0}^{(w)}(t_0) \). We have for any \( t_0 \)

\[ \int_{-\infty}^{\infty} S_{t_0}^{(w)}(t_0, f) \, df = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x(r) w(r-t) e^{-2 \pi ifr} \, dr \, df \]  (2.4)

Here we have used Parseval's formula [1, p. 65]

\[ \int_{-\infty}^{\infty} |Y(f)|^2 \, df = \int_{-\infty}^{\infty} |y(t)|^2 \, dt \]  (2.5)

with \( y(t) = x(t) w(t) \). Similarly, by using (1.2), we have for any \( f_0 \)

\[ \int_{-\infty}^{\infty} S_{t_0}^{(w)}(t, f_0) \, dt = \int_{-\infty}^{\infty} |X(g) W(f-g)|^2 \, dg \]  (2.6)

Inserting (2.4) and (2.6) into (2.2) and (2.3), we get after a simple manipulation

\[ \Sigma_{t_0}^{(w)}(t_0) = \int_{-\infty}^{\infty} (r - t_0)^2 |X(r)|^2 \, dr + \int_{-\infty}^{\infty} r^2 |W(f)|^2 \, df \]  (2.7)

\[ \Sigma_{t_0}^{(w)}(f_0) = \int_{-\infty}^{\infty} (f - f_0)^2 |X(f)|^2 \, df + \int_{-\infty}^{\infty} f^2 |W(f)|^2 \, df \]  (2.8)

Here we have used (1.5) and (1.7), and the fact that

\[ \int_{-\infty}^{\infty} (r - t_0)^2 |X(r)|^2 \, dr = \int_{-\infty}^{\infty} (f - f_0)^2 |X(f)|^2 \, df = 0 \]  (2.9)

Adding (2.7) and (2.8) we readily obtain (1.10).

We next show that \( w(t) \) of (1.8) uniquely minimizes \( \sigma_{t_0}^2(0, 0) \). We have for any \( w(t) \)

\[ \sigma_{t_0}^2(0, 0) = \int_{-\infty}^{\infty} r^2 |w(t)|^2 \, dt + \int_{-\infty}^{\infty} f^2 |W(f)|^2 \, df \]

\[ \geq 2 \left( \int_{-\infty}^{\infty} r^2 |w(t)|^2 \, dt \right)^{1/2} \left( \int_{-\infty}^{\infty} f^2 |W(f)|^2 \, df \right)^{1/2} \]

\[ \geq \frac{1}{2\pi} \]  (2.10)

Formulas (2.4) and (2.6) were presented in [2, eq. (4.8) and (4.9)], without proof.
In the first inequality in (2.10) (which is the elementary inequality $a^2 + b^2 \geq 2ab$, $a > 0$, $b > 0$) we have equality if and only if 
$$\sum_{m=1}^{n} t^2 |w(t)|^2 dt = \sum_{m=1}^{n} f^2 |W(f)|^2 df.$$ (2.11)

In the second inequality in (2.10) (which is the classical Heisenberg inequality, see [1, p. 273]) we have equality if and only if $w(t)$ is of the form $c(t^j)^d \exp (-\alpha t^2)$ for some $\alpha > 0$ and some complex $c$ with $|c| = 1$. Inserting this special form into (2.11), we readily obtain $\alpha = 1$, and the proof is complete.

ACKNOWLEDGMENT

The author is indebted to both referees who contributed considerably to a better presentation and who suggested important improvements and shortcuts in the author’s original proof.

REFERENCES


Comparison of Algorithms for Standard Median Filtering

Martti Juhola, Jyrki Katajainen, and Timo Raita

Abstract—In standard median filtering we search repeatedly for a median from a sample set which changes only slightly between the subsequent searches. We review several well-known methods for solving this running median problem, analyze the (asymptotical) time complexities of the methods, and propose simple variants which are especially suited for small sample sets, a frequent situation. Although we have restricted our discussion to the one-dimensional case, the ideas are easily extended to higher dimensions.

I. INTRODUCTION

A signal is here a sequence $S = (x_1, x_2, \cdots, x_k)$ of discrete values from a finite range. In median filtering [1]–[8] we slide a window of $m = 2k + 1$ samples and output 

$$y_i = \text{median}(x_i|j = i - k, \cdots, i + k).$$

The value $y_{i+k}$ is determined by rejecting the oldest sample $x_{i-k}$, inserting $x_{i+k+1}$, and searching the median again. We study data structures and algorithms for this running median problem. The view we have chosen rules out hardware-oriented techniques [6], [9], and [10], which are based on the bit representation of samples. Most methods can be extended to higher dimensions by replacing each sample by a vector of samples. We devise a data structure $W$ for the set $S$ supporting the following operations:

create($S$) returns $W$ containing $x_1, x_2, \cdots, x_{2k+1}$.
findmedian($W$) returns the median of the set of samples in $W$.
replace($x$, $y$, $W$) removes the item $y$ from $W$ and adds $x$ into $W$.
new($t$) returns the next sample value
oldest($t$) returns the sample added to the set at time $t - k$.

Create is used only once in the initialization. The two others are applied once at each step. Therefore the total time complexity of the latter ones is important. Replace is time dependent and could be accomplished by:

insert($x$, $W$) adds a new sample item $x$ into $W$ and delete($y$, $W$) removes the item $y$ from $W$.

These operations imply that we should maintain the samples within a queue [11]–[14]. Often, however, the data structure which supports the time dependent operations is indistinguishable from the "main" structure $W$.

We can find efficient data structures for time independent findmedian, insert, and delete operations. For heaps [15] or search trees [14], each of them requires only $O(\log n)$ time. Thus they are theoretically optimal as to the asymptotic time complexity. Applying a sequence of $3k + 2$ (findmedian, replace) pairs we can sort a set $(x_1, \cdots, x_{k+1})$ as follows. Let $-\infty (+\infty)$ denote a dummy element smaller (respectively, larger) than any of $x_i$. Using these dummy elements we can sort the elements $x_i$ ($i = 1, \cdots, k+1$) by applying findmedian and replace operations. First, we form a set of $2k+1$ elements by performing $k$ (findmedian($W$), replace ($-\infty$, $W$)) operation pairs followed by $k + 1$ (findmedian($W$), replace($x_i$, $W$)) pairs ($i = 1, \cdots, k+1$) where the findmedians can be considered as empty. After this the $k + 1$ (findmedian($W$), replace($+\infty$, $W$)) operations give the elements $x_i$ in sorted order.

By surrounding each block of $k+1$ samples of the original signal with the dummy elements as described, we have a sequence the length of which is still $O(n)$. For sorting $k$ items we have an information theoretic lower bound $\Omega(k \log k)$. Thus, we have a lower bound $\Omega(n \log m)$ for filtering a signal of length $O(n)$ with a window size $m$.

We shall present five ways to organize the data for efficient median filtering. We start with a structure in which no dependencies between the samples are stored. Then we gradually bring in facilities for characterizing more of the dependencies. In consequence, replace becomes slower and findmedian faster as we advance. The first two methods do not use sorting techniques in order to accomplish the median search [5], the next two methods do. In the last section we take advantage of the fact that the samples are integers from a given range.

All algorithms need a structure which holds the samples inside the window in time order ("age"). We use an array of length $m$ with a time pointer indicating the current oldest value. This is called a time ring (the oldest value is replaced at each step by a new value).

The time independent operations are performed in structure $W$. Each slot of the ring contains a pointer to the corresponding item in $W$.