Proceedings of the Federated Conference on Computer Science and Information Systems pp. 673–678
DOI: 10.15439/2016F455
ACSIS, Vol. 8. ISSN 2300-5963

Error analysis for the first-order Gaussian recursive filter operator

Ardelio Galletti, Giulio Giunta
University of Naples “Parthenope”
Department of Science and Technology
Centro Direzionale, Isola C4, 80143, Naples, Italy
Email: {ardelio.galletti, giulio.giunta}@uniparthenope.it

***Abstract*—Nowadays, recursive filters (RFs) are frequently used in several research fields. More in particular, Gaussian RFs offer a more efficient way for computing approximate Gaussian filters and Gaussian-based convolutions. The use of such recursive filters introduces many sources of errors. Among them, here we consider the filter truncation error, that is the error due to the transition from the starting filter operator to the RF approximating it. Since input and output signals have infinite dimensions, the analysis of the related filter operator involves infinite matrices. In this paper, starting from a summary of the comprehensive mathematical background, we consider the case of the first-order Gaussian recursive filter. Then, taking into account the matrix form of the related operator, we perform the error analysis and provide theoretical results that estimate the filter truncation error.**

I. Introduction

IN RECENT years, recursive filters have been frequently used in several fields. For example Gaussian RFs are usually involved in image processing [1], [2] and are also implemented for solving three-dimensional variational analysis schemes in data assimilation [3]. Moreover, they have been recently constructed specifically for the electrocardiogram denoising [4], [5], [6]. The idea of recursive filters is to approximate a given filter, or for example the convolution with the impulse response of such a filter, in a more efficient way. More in particular, among RFs, the Gaussian RFs are very efficient implementations that approximate Gaussian-based convolutions. Gaussian RFs can be constructed in several ways but, in this work, we deal just with the kind derived by Deriche [7] and Young and van Vliet (see [8], and the references therein). As is well known, Gaussian RF based algorithms, applied to a signal with compact support, generate unbounded distortions in the output signal boundary entries (a detailed explanation is in [8]). This is known as edge effect and, to the aim of removing it, theoretical tools (named boundary conditions) and implementative improvements have been proposed in literature [8], [9]. Here we are not interested in edge effects and only focus on the error due to the the transition from the starting filter operator to the RF approximating it. We refer to this error as the filter truncation error. In this work, we are interested in studying the filter truncation error for the case of the first-order Gaussian recursive filter. The aim is to investigate on the quality of the approximation supplied by that filter. We underline that input and output signals have infinite dimensions, hence the analysis of the related filter operator will involve infinite matrices.

The paper is organized as follows. In Section 2, we give some mathematical preliminaries about the first-order Gaussian RF and also provide its matrix formulation. In Section 3, the analysis of the filter operator structure is carried out. In Section 4, we report the error analysis and provide an upper bound for the filter truncation error. Finally, conclusions in Section 5 close the paper.

II. Mathematical background

Let:

$$s^{(0)} = \left\{s_j^{(0)}\right\}_{j\in\mathbf{Z}} = \left(\ldots, s_{-2}^{(0)}, s_{-1}^{(0)}, s_0^{(0)}, s_1^{(0)}, s_2^{(0)}, \ldots\right)$$

be a input signal. $s^{(0)}$ can be thought of as a complex function defined on the set of integers, that is an element of the set of sequences of complex numbers $\mathbf{C}^{\mathbf{Z}}$. Let g denote the Gaussian function with zero mean and standard deviation σ. Let:

$$\delta_j = \begin{cases} 1 & \text{if } j = 0 \\ 0 & \text{if } j \neq 0 \end{cases} \tag{1}$$

be the *unit-sample*. The Gaussian filter is a filter whose impulse response to the unit-sample, i.e. the output of such a filter when the input is δ, is the Gaussian function g, or an approximation to it. Applying the Gaussian filter to the input $s^{(0)}$ gives rise to a response that can be simply expressed by the discrete Gaussian convolution:

$$s_j^{(g)} = \left(g * s^{(0)}\right)_j = \sum_{t=-\infty}^{+\infty} g_{j-t} s_t^{(0)}, \qquad \forall j \in \mathbf{Z}, \tag{2}$$

where:

$$g_t \equiv g(t) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{t^2}{2\sigma^2}\right). \tag{3}$$

The expression in (2) can be conveniently rewritten by changing the index t of the summation in $t - j$, and by making use of the symmetry $g_t = g_{-t}$. We have:

$$s_j^{(g)} = \sum_{t=-\infty}^{+\infty} g_t s_{j+t}^{(0)}, \qquad \forall j \in \mathbf{Z}. \tag{4}$$

The entries $s_j^{(g)}$ of $s^{(g)}$ can be efficiently approximated by means of Gaussian RFs and K-iterated Gaussian RFs.

A K-iterated n-order Gaussian RF filter computes the output

978-83-60810-90-3/$25.00©2016, IEEE

signal $s^{(K)}$, i.e. the K-iterate approximation of $s^{(g)}$, whose entries solve the infinite sequences of equations:

$$p_j^{(k)} = \beta s_j^{(k-1)} + \sum_{t=1}^{n} \alpha_t p_{j-t}^{(k)}, \qquad \forall\, j \in \mathbf{Z}, \tag{5}$$

$$s_j^{(k)} = \beta p_j^{(k)} + \sum_{t=1}^{n} \alpha_t s_{j+t}^{(k)}, \qquad \forall\, j \in \mathbf{Z}. \tag{6}$$

The filter iteration counter k goes from 1 to K (number of filter iterations). For $K = 1$, the filter merely becomes an n-order Gaussian RF filter. Equations in (5) and (6) are conveniently referred to as the advancing and backing filters, respectively: when a Gaussian RF is implemented as an algorithm, the index j must be treated in increasing order in the former and decreasing in the latter [8]. The values α_t and β are called *smoothing coefficients* and satisfy the constraint:

$$\beta = 1 - \sum_{t=1}^{n} \alpha_t.$$

In a general setting they depend on σ, n and K. In the following we consider just the first-order Gaussian RF with one-iteration only ($n = 1, K = 1$), for which equations (5) and (6) take the simplified form:

$$p_j = \beta s_j^{(0)} + \alpha p_{j-1}, \qquad \forall\, j \in \mathbf{Z}, \tag{7}$$

$$s_j = \beta p_j + \alpha s_{j+1}, \qquad \forall\, j \in \mathbf{Z}. \tag{8}$$

The smoothing coefficients are given by:

$$\alpha = 1 + E_\sigma - \sqrt{E_\sigma(E_\sigma + 2)} \tag{9}$$

and:

$$\beta = \sqrt{E_\sigma(E_\sigma + 2)} - E_\sigma, \tag{10}$$

with $E_\sigma = \sigma^{-2}$. The behaviour of α and β, as σ varies, is shown in Figure 1. Using Taylor expansion arguments, we can observe that, for small σ, it is:

$$\alpha = \frac{1}{2}\sigma^2 - \sigma^4 + \mathcal{O}(\sigma^6), \quad (\text{for} \quad \sigma \to 0),$$

while, for large σ, it is:

$$\alpha = 1 - \sigma^{-1}\sqrt{2} + \mathcal{O}(\sigma^{-2}), \quad (\text{for} \quad \sigma \to +\infty).$$

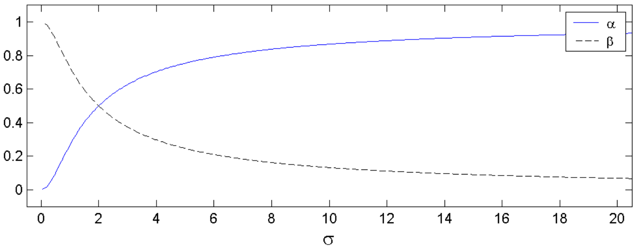

Fig. 1. Blue solid line: α. Black dashed line: β

By adapting the discussion in [3] to the case of the infinite dimension signals:

$$s^{(0)} = \{s_j^{(0)}\}_{j \in \mathbf{Z}}, \quad p = \{p_j\}_{j \in \mathbf{Z}} \quad \text{and} \quad s = \{s_j\}_{j \in \mathbf{Z}},$$

we can rewrite, in matrix form, the advancing filter (7) as:

$$Lp = s^{(0)}, \tag{11}$$

and the backing filter (8) as:

$$Us = p. \tag{12}$$

L and U are (bi)infinite matrices (matrices with infinite rows and columns, see [10] for notation and conventions), whose nonzero entries are:

$$L_{i,i} = \frac{1}{\beta}, \qquad L_{i,i-1} = -\frac{\alpha}{\beta}, \qquad \forall\, i \in \mathbf{Z}, \tag{13}$$

$$U_{i,i} = \frac{1}{\beta}, \qquad U_{i,i+1} = -\frac{\alpha}{\beta}, \qquad \forall\, i \in \mathbf{Z}. \tag{14}$$

Moreover L and U are both Toeplitz, bidiagonal matrices and U is the transpose of L, in the sense that:

$$U_{i,j} = L_{j,i}, \qquad \forall\, i,j \in \mathbf{Z}. \tag{15}$$

The products in (11) and (12) can be thought of as multiplications of infinite matrices. Given two infinite matrices:

$$A = \{A_{i,j}\}_{i,j \in \mathbf{Z}} \quad \text{and} \quad B = \{B_{i,j}\}_{i,j \in \mathbf{Z}},$$

the matrix product $C = AB$ has entries expressed as the series:

$$C_{i,j} = A_i B^j = \sum_{k=-\infty}^{+\infty} A_{i,k} B_{k,j},$$

where A_i and B^j are sequences denoting a row of A and a column of B, respectively. Now, substituting in equation (11) the expression of p given in (12), we deduce that the output s and the input $s^{(0)}$ of the first-order Gaussian RF satisfy:

$$Ds = s^{(0)}, \quad \text{with} \quad D = LU. \tag{16}$$

Then, to obtain the operator form for such a filter, we need just to invert the matrix D in (16). For infinite matrices, many definitions of inverse are given. Here, we mean the matrix A^{-1} as the inverse of an infinite matrix A, if and only if:

$$AA^{-1} = A^{-1}A = I,$$

where $I_{i,j} = \delta_{i-j}$, with δ_j as in (1). If D has inverse D^{-1}, from (16) it trivially results:

$$s = D^{-1}s^{(0)}. \tag{17}$$

Equation (17) proves that the infinite matrix:

$$F \overset{\text{def}}{=} D^{-1} = (LU)^{-1} \tag{18}$$

acts as the operator related to first-order Gaussian recursive filter. In the next section we will provide the structure of D and F.

III. Filter operator

We need the following results: the first lemma provides a formula, equivalent to (7) and (8), which expresses the output entries s_j in terms of the input entries $s_j^{(0)}$, without using p_j values; the second lemma makes explicit the expression of the entries of the infinite matrix product LU.

Lemma III.1. *(Filter output entries representation) Let s, p_j and $s_j^{(0)}$ be as in* (7) *and* (8). *If $\beta = 1-\alpha$ and $|\alpha| < 1$, then the output entries s_j are given by the series:*

$$s_j = \sum_{t=-\infty}^{+\infty} c_t s_{j+t}^{(0)}, \qquad \forall j \in \mathbf{Z}, \tag{19}$$

with:

$$c_t \equiv \frac{\beta}{1+\alpha}\alpha^{|t|}, \qquad \forall t \in \mathbf{Z}. \tag{20}$$

Proof. Let k be a positive integer. Combining the equation (7) in itself repeatedly, with indices j, $j-1, \ldots, j-(k-1)$, we obtain inductively:

$$p_j = \beta \sum_{m=0}^{k-1} \alpha^m s_{j-m}^{(0)} + \alpha^k p_{j-k}, \qquad \forall j \in \mathbf{Z},$$

and, since $|\alpha| < 1$, for $k \to +\infty$ it is:

$$p_j = \beta \sum_{m=0}^{+\infty} \alpha^m s_{j-m}^{(0)}, \qquad \forall j \in \mathbf{Z}, \tag{21}$$

Similarly, combining the equation (8) in itself, with indices j, $j+1, \ldots, j+(k-1)$, we get:

$$s_j = \beta \sum_{l=0}^{k-1} \alpha^l p_{j+l} + \alpha^k s_{j+k}, \qquad \forall j \in \mathbf{Z},$$

and, for $k \to +\infty$ it is:

$$s_j = \beta \sum_{l=0}^{+\infty} \alpha^l p_{j+l}, \qquad \forall j \in \mathbf{Z}. \tag{22}$$

Hence, using (21) with $j+l$ in (22), we obtain:

$$\begin{aligned} s_j &= \beta \sum_{l=0}^{+\infty} \alpha^l \left(\beta \sum_{m=0}^{+\infty} \alpha^m s_{j+l-m}^{(0)} \right) \\ &= \sum_{l=0}^{+\infty} \sum_{m=0}^{+\infty} \beta^2 \alpha^{l+m} s_{j+l-m}^{(0)}, \qquad \forall j \in \mathbf{Z}. \end{aligned} \tag{23}$$

Observing that, as l, m vary in $0, 1, \ldots$, $t = l - m$ varies in $\mathbf{Z}$, we can simplify the representation in (23) by collecting the coefficients of $s_{j+l-m}^{(0)} = s_{j+t}^{(0)}$, for each fixed t. Then, by means of a suitable change of indices in (23), we get:

$$s_j = \sum_{t=-\infty}^{+\infty} \left(\sum_{m=0}^{+\infty} \beta^2 \alpha^{|t|+2m} \right) s_{j+t}^{(0)}, \qquad \forall j \in \mathbf{Z}.$$

Finally, the thesis follows using $\beta = 1-\alpha$ and because of:

$$\sum_{m=0}^{+\infty} \beta^2 \alpha^{|t|+2m} = \beta^2 \alpha^{|t|} \sum_{m=0}^{+\infty} \alpha^{2m} = \frac{\beta^2}{1-\alpha^2} \alpha^{|t|} = c_t.$$

□

Lemma III.2. *(D structure) Let L and U be with entries as in* (13) *and* (14), *respectively. Then $D = LU$ is a Toeplitz, tridiagonal, symmetric, infinite matrix, with entries:*

$$D_{i,j} = \begin{cases} \frac{1+\alpha^2}{\beta^2} & \textit{if } j = i \\ -\frac{\alpha}{\beta^2} & \textit{if } j = i \pm 1 \\ 0 & \textit{if } j = i \pm m, \quad m \geq 2 \end{cases} \tag{24}$$

Proof. We need just to prove (24). Since $L_{i,k} = 0$ for $k < i-1$ and $k > i$, $\forall i, j \in \mathbf{Z}$ it is:

$$D_{i,j} = \sum_{k=-\infty}^{+\infty} L_{i,k} U_{k,j} = L_{i,i-1} U_{i-1,j} + L_{i,i} U_{i,j}.$$

Then recalling (13) and (15), it holds:

$$D_{i,j} = L_{i,i-1} L_{j,i-1} + L_{i,i} L_{j,i} = -\frac{\alpha}{\beta} L_{j,i-1} + \frac{1}{\beta} L_{j,i}. \tag{25}$$

Putting $j = i$ in (25), we obtain:

$$D_{i,i} = -\frac{\alpha}{\beta} L_{i,i-1} + \frac{1}{\beta} L_{i,i} = \left(-\frac{\alpha}{\beta}\right)^2 + \left(\frac{1}{\beta}\right)^2 = \frac{1+\alpha^2}{\beta^2}.$$

For $j = i-1$, it is:

$$D_{i,i-1} = -\frac{\alpha}{\beta} L_{i-1,i-1} + \frac{1}{\beta} L_{i-1,i} - \frac{\alpha}{\beta}\frac{1}{\beta} + \frac{1}{\beta} \cdot 0 = -\frac{\alpha}{\beta^2},$$

and for $j = i+1$, it is:

$$D_{i,i+1} = -\frac{\alpha}{\beta} L_{i+1,i-1} + \frac{1}{\beta} L_{i+1,i} = \frac{1}{\beta}\left(-\frac{\alpha}{\beta}\right) = -\frac{\alpha}{\beta^2}.$$

Finally, for $j = i \pm m$, with $m \geq 2$, we have:

$$L_{j,i-1} = L_{i \pm m, i-1} = 0$$

and:

$$L_{j,i} = L_{i \pm m, i} = 0.$$

Hence, from (25), we obtain:

$$D_{i,j} = D_{i,i+m} = L_{i,i-1} L_{i \pm m, i-1} + L_{i,i} L_{i \pm m, i} = 0,$$

and this completes the proof. □

Using the result in Lemma III.1 we can derive the structure of the operator $F = D^{-1}$. Notice that, from (17) and (18), it is:

$$s_j = \left(F s^{(0)}\right)_j = F_j s^{(0)} = \sum_{k=-\infty}^{+\infty} F_{j,k} s_k^{(0)}.$$

With the substitution $k = j + t$ we obtain the expression:

$$s_j = \sum_{t=-\infty}^{+\infty} F_{j,j+t} s_{j+t}^{(0)}, \tag{26}$$

which has the same form of the result in (19). This suggests that the F entries are given by the coefficients in (20). Starting from this remark, we deduce the form of F.

Theorem III.1. *(F structure) Let L and U be with entries as in* (13) *and* (14), *respectively. Then $F = (LU)^{-1}$ is a Toeplitz, symmetric, infinite matrix, with entries:*

$$F_{i,j} = \frac{\beta}{1+\alpha} \alpha^{|j-i|}, \qquad \forall i, j \in \mathbf{Z}. \tag{27}$$

Proof. By comparing the series in (26) and (19), for each fixed $j \in \mathbf{Z}$, we get:

$$\sum_{t=-\infty}^{+\infty} F_{j,j+t} s^{(0)}_{j+t} = \sum_{t=-\infty}^{+\infty} c_t s^{(0)}_{j+t}.$$

Therefore, by taking for all $t \in \mathbf{Z}$, the input signal $s^{(0)}$ as the time shifted unit-sample with nonzero entry $s^{(0)}_{j+t}$, it follows:

$$F_{j,j+t} = c_t, \qquad \forall\, t \in \mathbf{Z}.$$

Then, recalling (20) we obtain the thesis:

$$F_{i,j} = F_{i,i+(j-i)} = c_{j-i} = \frac{\beta}{1+\alpha}\alpha^{|j-i|}, \qquad \forall\, i,j \in \mathbf{Z}.$$

□

Another proof that F actually acts as the inverse of D, can be obtained by verifying, by direct computation, that $DF = I$. To do this, observe that from Lemma III.2, we have:

$$(DF)_{i,j} = \sum_{k=-\infty}^{+\infty} D_{i,k}F_{k,j} = \sum_{k=i-1}^{i+1} D_{i,k}F_{k,j}.$$

So, for $j = i + m$, it is:

$$(DF)_{i,i+m} = D_{i,i-1}F_{i-i,i+m} + D_{i,i}F_{i,i+m} + D_{i,i+1}F_{i+1,i+m},$$

and by exploiting (24) and (27), we get:

$$(DF)_{i,i+m} = \frac{-\alpha\alpha^{|m+1|} + (1+\alpha^2)\alpha^{|m|} - \alpha\alpha^{|m-1|}}{\beta(1+\alpha)}. \quad (28)$$

For $j = i$, that is $m = 0$, (28) becomes:

$$(DF)_{i,i} = \frac{-\alpha^2 + (1+\alpha^2) - \alpha^2}{\beta(1+\alpha)} = \frac{1-\alpha^2}{\beta(1+\alpha)} = 1.$$

For $j > i$, that is $m \geq 1$, (28) becomes:

$$\begin{aligned}(DF)_{i,i+m} &= \frac{-\alpha\alpha^{m+1} + (1+\alpha^2)\alpha^m - \alpha\alpha^{m-1}}{\beta(1+\alpha)} \\ &= \alpha^m \frac{-\alpha^2 + (1+\alpha^2) - 1}{\beta(1+\alpha)} = 0.\end{aligned}$$

For $j < i$, that is $m \leq -1$, (28) becomes:

$$\begin{aligned}(DF)_{i,i+m} &= \frac{-\alpha\alpha^{-m-1} + (1+\alpha^2)\alpha^{-m} - \alpha\alpha^{-m+1}}{\beta(1+\alpha)} \\ &= \alpha^{-m} \frac{-1 + (1+\alpha^2) - \alpha^2}{\beta(1+\alpha)} = 0.\end{aligned}$$

IV. Error analysis

In this section we are interested in studying the error occurring when the Gaussian filter is substituted by the first-order Gaussian RF, namely the filter truncation error. Let $\|\cdot\|$ denote the sup-norm, defined for signals f as:

$$\|f\| = \sup_{k\in\mathbf{Z}} |f_k|,$$

and for infinite matrices A as:

$$\|A\| = \sup_{f\in\mathbf{C}^{\mathbf{Z}}, \|f\|=1} \|Af\| = \sup_{i\in\mathbf{Z}} \sum_{j\in\mathbf{Z}} |A_{i,j}|.$$

Let denote by:

$$\tau_j = s^{(g)}_j - s_j, \quad (29)$$

the difference between the output entries of the Gaussian filter and the first-order Gaussian RF. We refer to:

$$\tau = \left\| \left\{\tau_j\right\}_{j\in\mathbf{Z}} \right\| \quad (30)$$

as the filter truncation error (f.t.e.). Before giving an upper bound for τ, let us indicate by V the infinite Gaussian matrix, with entries:

$$V_{i,j} = g_{j-i}, \qquad \forall\, i,j \in \mathbf{Z}, \quad (31)$$

where g_t values are as in (3). Now, using the operator V, the equation (4) is compactly represented as:

$$s^{(g)} = Vs^{(0)}. \quad (32)$$

Therefore, combining (17), (18), (29) and (32), we get:

$$\left\{\tau_j\right\}_{j\in\mathbf{Z}} = s^{(g)} - s = Vs^{(0)} - Fs^{(0)} = (V-F)s^{(0)}, \quad (33)$$

that is the f.t.e. is simply obtained as the product of the filter operators difference and the input signal. Starting from (33) we can proof the following main result.

Theorem IV.1. *(Filter truncation error) Assume that $\|s^{(0)}\| \leq S$ and let V be the same as in* (31)*. Then, for the f.t.e. defined in* (29) *and* (30)*, it holds that:*

$$\tau \leq \kappa \cdot S, \quad (34)$$

with:

$$\kappa = \|V - F\| = \sum_{t=-\infty}^{+\infty} |g_t - c_t| \quad (35)$$

and g_t and c_t as in (3) *and* (20)*, respectively.*

Proof. Immediate by construction. The proof of (34) is as follows. From (30) and (33) it is:

$$\tau = \|(V-F)s^{(0)}\| \leq \|V-F\| \cdot \|s^{(0)}\| = \kappa \cdot S.$$

To complete the proof, we need just to prove (35). From (20) and (27) we deduce that $\forall\, i,t \in \mathbf{Z}$ it is:

$$F_{i,i+t} = \frac{\beta}{1+\alpha}\alpha^{|(i+t)-i|} = \frac{\beta}{1+\alpha}\alpha^{|t|} = c_t.$$

Then, changing the summation index j in $i+t$, and by using (31), we get the thesis:

$$\begin{aligned}\kappa = \|V - F\| &= \sup_{i\in\mathbf{Z}} \sum_{j\in\mathbf{Z}} |V_{i,j} - F_{i,j}| \\ &= \sup_{i\in\mathbf{Z}} \sum_{t\in\mathbf{Z}} |V_{i,i+t} - F_{i,i+t}| = \sup_{i\in\mathbf{Z}} \sum_{t\in\mathbf{Z}} |g_t - c_t| \\ &= \sum_{t\in\mathbf{Z}} |g_t - c_t| = \sum_{t=-\infty}^{+\infty} |g_t - c_t|. \quad (36)\end{aligned}$$

□

The previous result proves that, without considering the order of magnitude S of the input signal, the factor κ behaves like a physical limit in the accuracy provided by the first-order Gaussian RF in approximating the Gaussian convolution. Then, for investigating on the f.t.e., and completing the error analysis, we can limit our discussion to the behaviour of κ. Recalling (3), (9), (10) and (20), we deduce that coefficients g_t and c_t depend on σ and t, making κ dependent just on σ. We remark that if $g_t - c_t$ was of constant sign, for example $g_t - c_t \geq 0$, $\forall\, t \in \mathbf{Z}$, then we would simply have:

$$\kappa = \sum_{t=-\infty}^{+\infty} g_t - \sum_{t=-\infty}^{+\infty} c_t = \sum_{t=-\infty}^{+\infty} g_t - 1 \leq \frac{1}{\sigma\sqrt{2\pi}}, \qquad (37)$$

where the last inequality arises from:

$$\sum_{t=-\infty}^{+\infty} g_t \leq 1 + \frac{1}{\sigma\sqrt{2\pi}}.$$

This bound can be easily proved exploiting the monotonicity properties of the Gaussian function:

$$g_t \leq \int_{t-1}^{t} g(x)dx, \qquad \forall t = 1, 2, \ldots,$$

and the symmetry $g_t = g_{-t}$. Indeed, we have:

$$\sum_{t=-\infty}^{+\infty} g_t = g_0 + 2\sum_{t=1}^{+\infty} g_t \leq \frac{1}{\sigma\sqrt{2\pi}} + 2\sum_{t=1}^{+\infty}\int_{t-1}^{t} g(x)dx$$
$$= \frac{1}{\sigma\sqrt{2\pi}} + 2\int_{0}^{+\infty} g(x)dx = \frac{1}{\sigma\sqrt{2\pi}} + 1.$$

However, in general, the coefficients $g_t - c_t$ change sign as t varies. An example of this fact is in Figure 2, where $g_t - c_t$ values are obtained for $\sigma = 100$. Consequently, (37) cannot be proved and κ is not bounded by $1/(\sigma\sqrt{2\pi})$. In fact, the behaviour of κ, as σ varies in $[0.05, 50\,000]$, is shown in Figure 3. The figure highlights that κ takes its minimum values $\kappa_{min} = 0.05$ for $\sigma \approx 0.47$ and, except for values of σ in a small interval ($[0.37, 0.60]$), is always greater than 0.25. Moreover, we observe the following asymptotic behaviours:

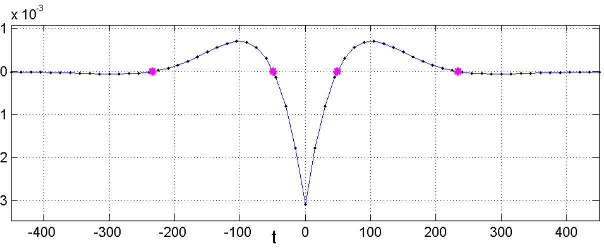

Fig. 2. Behaviour of $g_t - c_t$ for $\sigma = 100$ and $t = -450, \ldots, 450$

- for small values of σ, κ seems to be unbounded and to increase like $1/\sigma$. This trend is consistent with (37) and is easily proved. Observing that $\alpha > 0$ implies:

  $$c_0 = \frac{\beta}{1+\alpha} = \frac{1-\alpha}{1+\alpha} < 1,$$

  for $\sigma < 1/\sqrt{2\pi} = 0.398$, it is:

  $$\frac{1}{\sigma\sqrt{2\pi}} - 1 \leq g_0 - c_0 = |g_0 - c_0| \leq \kappa;$$

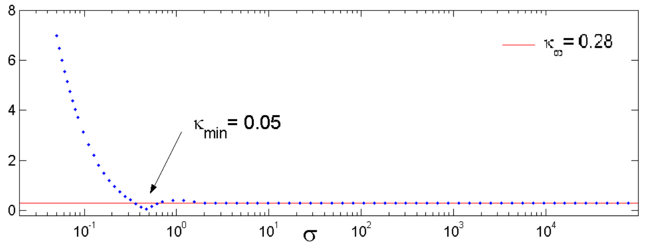

Fig. 3. Behaviour of κ for 75 values of σ increasing exponentially in the interval $[0.05, 50\,000]$

- for σ large enough, ($\sigma > 3.4$), κ becomes indeed constant and takes nearly the asymptotic value:

  $$\kappa_\infty = \lim_{\sigma\to\infty} \kappa \approx 0.28.$$

This is the empirical evidence that (37) does not hold true for all σ. The value κ_∞ can be found observing that, for large σ, κ can be accurately approximated by the integral:

$$\int_{-\infty}^{+\infty} |g(t) - c(t)|dt,$$

where g is the Gaussian function, and c is the function:

$$c(t) = \frac{\beta}{1+\alpha}\alpha^{|t|}, \qquad \forall\, t \in \mathbf{R}.$$

To compute the integral, we need to establish when $g - c$ changes sign. Figure 2 shows that, for $\sigma = 100$, $g - c$ has 4 zeros (two pairs: $\pm t_1$, $\pm t_2$) and that changes sign 4 times. That is what actually happens for each large enough σ. Indeed, using the approximations:

$$\alpha \approx 1 - \frac{\sqrt{2}}{\sigma},\ \beta \approx \frac{\sqrt{2}}{\sigma},\ \frac{1}{1+\alpha} \approx \frac{1}{2},\ 1 - \frac{\sqrt{2}}{\sigma} \approx \exp\left(-\frac{\sqrt{2}}{\sigma}\right),$$

we obtain:

$$c(t) \approx \frac{\sqrt{2}}{\sigma}\frac{1}{2}\left(1 - \frac{\sqrt{2}}{\sigma}\right)^{|t|} \approx \frac{1}{\sigma\sqrt{2}}\exp\left(-\sqrt{2}\frac{|t|}{\sigma}\right) = \tilde{c}(t).$$

Solving $g(t) = \tilde{c}(t)$ for $x = \frac{|t|}{\sigma}$, we get:

$$\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right) = \frac{1}{\sigma\sqrt{2}}\exp\left(-\sqrt{2}x\right),$$

from which, taking the logarithms:

$$-x^2 - \ln\pi = -2x\sqrt{2},$$

and so:

$$x_1 = \sqrt{2} - \sqrt{2 - \ln\pi} = 0.489, \quad \text{and} \quad t_1 = x_1\sigma,$$

and:

$$x_2 = \sqrt{2} + \sqrt{2 - \ln\pi} = 2.339, \quad \text{and} \quad t_2 = x_2\sigma.$$

Finally, the value κ_∞ is achieved exploiting the symmetry and the sign of $|c - g|$. We have:

$$\kappa_\infty = 2\int_0^{+\infty} |g(t) - \tilde{c}(t)|dt = 2\int_0^{x_1\sigma} \big(\tilde{c}(t) - g(t)\big)dt +$$
$$+2\int_{x_1\sigma}^{x_2\sigma} \big(g(t) - \tilde{c}(t)\big)dt + 2\int_{x_2\sigma}^{+\infty} \big(\tilde{c}(t) - g(t)\big)dt = 0.28,$$

where, after changing the variable of integration t in $x = t/\sigma$ and specifying computations, one can see that the values of the integrals are no longer dependent on σ. In conclusion, our analysis has pointed out that, except for a small subinterval of σ values, the bound κ of the filter truncation error is never significantly small. Then we can state that the first-order Gaussian RF, when used in a single iteration, does not offer a good approximation of the Gaussian convolution.

V. Conclusions

In this work, we have given mathematical preliminaries and definitions about Gaussian RFs by focusing on the first-order Gaussian RF. For this filter we have studied the related matrix operator, by providing a complete description of its structure. Then, we have studied the associated filter truncation error and, exploiting the structure operator, we have given a theoretical error upper bound. This result has been used to investigate about the quality of the approximation supplied by that filter and has allowed to conclude that, in general, this filter is not very accurate.

References

[1] van Vliet, L.J., Young, I.T., Verbeek, P.W.. - *Recursive Gaussian derivative filters*. The 14 th International Conference on Pattern Recognition, pp. 509-514, DOI: 10.1109/ICPR.1998.711192, 1998.

[2] Young, I.T., van Vliet L.J.. - *Recursive implementation of the Gaussian filter*. Signal Processing 44, pp 139-151, 1995.

[3] Cuomo, S., Farina, R., Galletti, A., Marcellino, L.. -*An error estimate of Gaussian recursive filter in 3Dvar problem* Federated Conference on Computer Science and Information Systems, FedCSIS 2014, art. no. 6933068, pp. 587-595, 2014. DOI: 10.15439/2014F279

[4] Cuomo, S., De Pietro, G., Farina, R., Galletti, A., Sannino, G.. - *A novel O(n) numerical scheme for ECG signal denoising* Procedia Computer Science, 51 (1), pp. 775-784, 2015. DOI: 10.1016/j.procs.2015.05.198

[5] Cuomo, S., De Pietro, G., Farina, R., Galletti, A., Sannino, G.. - *A framework for ECG denoising for mobile devices* PETRA 2015 ACM. ISBN 978-1-4503-3452-5/15/07, DOI: 10.1145/2769493.2769560, 2015.

[6] Cuomo, S., De Pietro, G., Farina, R., Galletti, A., Sannino, G.. - *A revised scheme for real time ECG Signal denoising based on recursive FILTERING*, Biomedical Signal Processing and Control, 27, pp. 134-144, 2016. DOI: 10.1016/j.bspc.2016.02.007

[7] R. Deriche - *Recursively implementating the Gaussian and its derivatives*. INRIA Research Report RR-1893, 1993, pp.24.

[8] Cuomo, S., Farina, R., Galletti, A., Marcellino, L.. - *A K-iterated scheme for the first-order Gaussian Recursive Filter with boundary conditions* Federated Conference on Computer Science and Information Systems, FedCSIS 2015, pp.641-647, 2015. DOI: 10.15439/2015F286

[9] Triggs, B., Sdika M.. - *Boundary conditions for Young-van Vliet recursive filtering*. IEEE Transactions on Signal Processing, 54 (6 I), pp. 2365-2367, 2006.

[10] de Boor, C., Jia, R.-q., Pinkus, A.. - *Structure of invertible (bi)infinite totally positive matrices* Linear Algebra and Its Applications, 47 (C), pp. 41-55, 1982. DOI: 10.1016/0024-3795(82)90225-7