

# Newton-Raphson Solution of Nonlinear Delay-Free Loop Filter Networks

Federico Fontana, *Senior Member, IEEE*, and Enrico Bozzo

## Abstract

For their numerical properties and speed of convergence, Newton methods are frequently used to compute nonlinear audio electronic circuit models in the digital domain. These methods are traditionally employed regardless of preliminary considerations about their applicability, primarily because of a lack of flexible mathematical tools making the convergence analysis an easy task. In this paper we derive a tool which is specific to the case when the nonlinear circuit can be modeled in terms of a delay-free loop network. In this case, a distance function can be defined from a known convergence theorem providing a sufficient condition for quadratic speed of convergence of the method. After substituting the nonlinear characteristics with equivalent linear filters which compute Newton-Raphson on the existing network, through this function we figure out constraints guaranteeing quadratic convergence speed in the diode clipper. Further application to a ring modulator circuit shows proportionality of the same function with the convergence speed in the resulting filter network. This case study suggests use of the proposed distance function as a speed predictor, with potential application in the design of virtual analogue systems for real-time digital audio effects.

## Index Terms

Digital delay-free loop, nonlinear filter network, Newton-Raphson method, diode clipper, ring modulator, virtual analogue.

## I. INTRODUCTION

A nonlinear delay-free loop (DFL from now on) is an elementary digital filter network containing no delay units along a nonlinear loopback [1], [2]. As Fig. 1(a) shows, there is no explicit

F. Fontana and E. Bozzo are with the Department of Mathematics, Computer Science and Physics, University of Udine, Udine, 33100 Italy e-mail: {federico.fontana,enrico.bozzo}@uniud.it.

Manuscript received April 19, 2005; revised August 26, 2015.

procedure allowing for the computation of the network. In fact, the discrete-time nonlinear block  $c(\cdot)$  propagates the output from the block instantaneously back to itself after summation with the input  $u$  to the network. In this way, at every temporal step  $n$  a nonlinear equation in the unknown  $v$  of the type  $v = c(v) + u[n]$  must be solved to find out the output  $v[n]$  from the network, in general requiring the use of a numerical method.

The DFL problem eventually appears during the discretization of a nonlinear circuit with feedback, in spite of the variables that are chosen for its solution: lumped (i.e. Kirchhoff) [3], wave [4], space-state [5] or transformed [6], [7]. The literature reports use of diverse numerical methods to compute it, for various applications: bisection for saturation filters [8]; fixed-point for the Dolby B decoder [9], the EMS VCS3 voltage-controlled filter [10], the ring modulator [11]; table lookups for nonlinear oscillators [12]; other lookup structures avoiding iterative solutions for vacuum tubes in amplifiers [13] and for sound synthesis of collisions [14]; linearization [15] or insertion of unit delays when stronger approximations are possible, e.g. in the Moog voltage-controlled filter [16].

Among such numerical methods, Newton-Raphson (NR from now on) is largely preferred for its speed of convergence and relatively simple implementation [17]. Application of this solver to digital audio has been reported in fret-string, mass-spring, friction models for musical instrument excitation [18], [19], [20], [21], in guitar amplifier, preamp and pedal simulations [22], [23], [24], [25], [26], [27], in physically-based piano strings [28], stick-membrane collisions [29], and more in general in lots of digital audio effects [9], [30], [31], [32], [33], [34]. With the continuing evolution of the modeling approaches toward robustness and efficiency, NR has not lost its appeal. In particular, recent development of Wave Digital Networks accounting for multiple nonlinearities [35], [36] has made Newton methods key in the solution of such networks, both multi-dimensional [37] and also one-dimensional in cases when the nonlinear elements of the network are scalar and they can be solved one by one in the hybrid Wave-Kirchhoff domain [38].

NR searches a root of a function  $f(v)$  by iterating the scheme

$$v_{[k+1]} = v_{[k]} - \frac{f(v_{[k]})}{f'(v_{[k]})}. \quad (1)$$

If applied to the network in Fig. 1(a), then, NR can be conveniently set to look for a zero of the function

$$f(v) = v - c(v) - u. \quad (2)$$

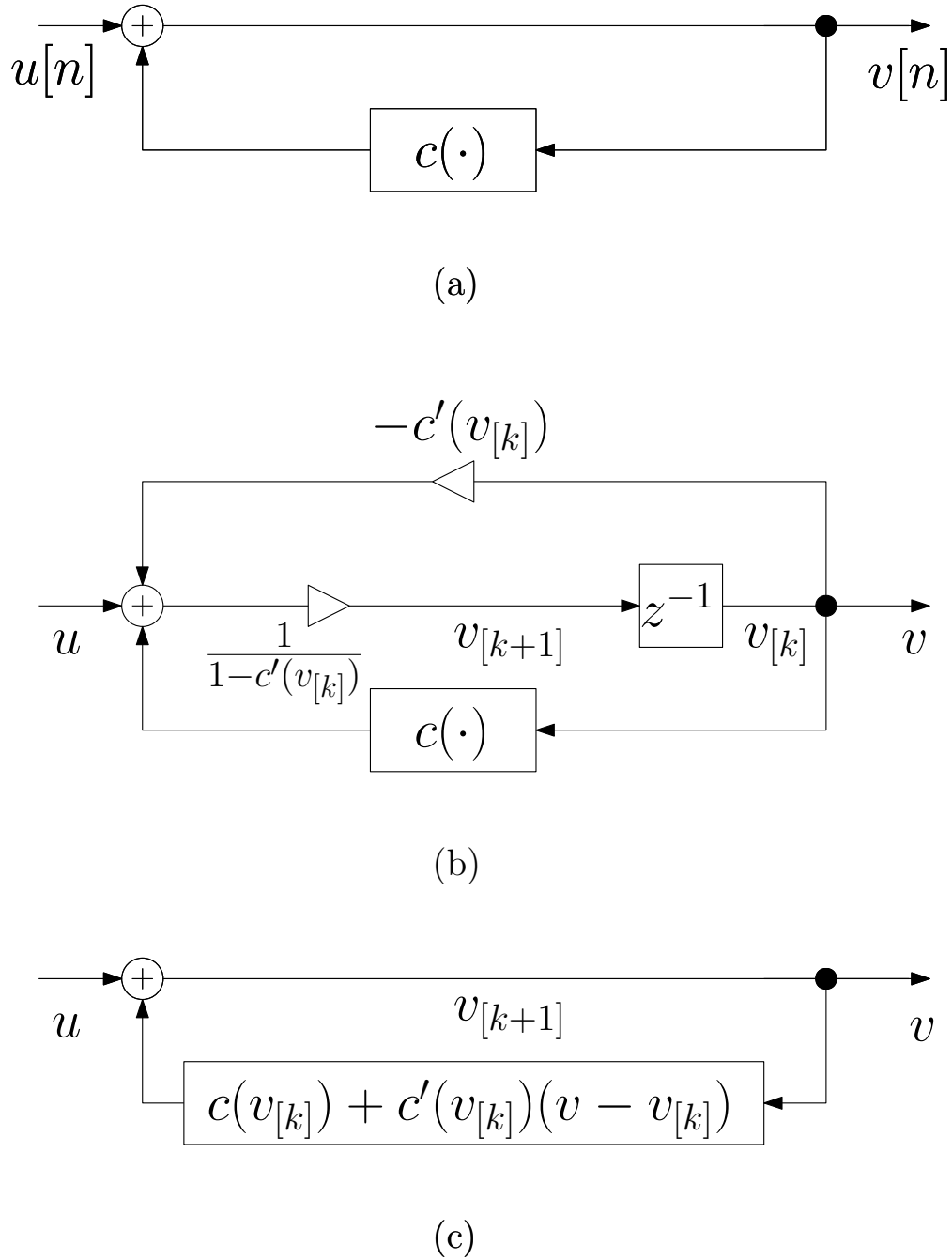


Fig. 1. Scalar nonlinear delay-free loop (a). Causal equivalent computing the output with NR (b). Linear equivalent at NR iteration step  $k$  (c).

In this case, (1) becomes

$$v_{[k+1]} = v_{[k]} - \frac{v_{[k]} - c(v_{[k]}) - u}{1 - c'(v_{[k]})} = \frac{c(v_{[k]}) - c'(v_{[k]})v_{[k]} + u}{1 - c'(v_{[k]})}. \quad (3)$$

Every iteration in this scheme is equivalent to a computation of the network in Fig. 1(b).

Borrowing terminology from [11], Fig. 1(b) then shows a *causal equivalent* of the network in Fig. 1(a) at temporal step  $n$ . It is straightforward to check that if NR converges then the causal equivalent reduces to the network in Fig. 1(a). In fact, if  $v_{[k+1]} = v_{[k]} = v$  then the upper loop in Fig. 1(b) provides

$$v = \frac{u - v c'(v)}{1 - c'(v)}, \quad (4)$$

which, after multiplication of both terms by  $1 - c'(v)$ , simplifies in  $v = u$ . At this point the upper loop has become uninfluent, hence it can be removed from the network along with the scaling factor and delay unit across the feed-forward branch.

Filter networks owe their success to the immediacy of access they offer to the signals flowing across the circuit blocks, not only in the lumped domain. Despite the maturity of the digital filter theory, they remain a versatile tool every time a processing system can be represented as an interconnection of scalar (including nonlinear) characteristics. This work continues research we have recently made on the computation of nonlinear DFL networks using fixed-point methods [11]. For these networks, in fact, a sufficient condition of convergence was found allowing for a straightforward solution that does not require to rearrange the network structure. Unfortunately, fixed-point solvers are known to be relatively slow especially if compared against methods whose speed of convergence is quadratic. On the other hand, Newton methods possess this property. Moved by the larger popularity of NR, this work investigates the existence of sufficient conditions for its fast (i.e., quadratic) convergence in nonlinear DFL networks, provided also the possibility to apply the method without disrupting the network structure.

In Sec. II it will be shown that NR computes a linear DFL equivalent preserving the structure of the nonlinear network. This equivalence will be applied to the diode clipper, a system which is simple and expressive enough to have frequently served as an example in the literature of nonlinear circuit models [32]—see [39, Sec. 2] for a list of references. Sec. III will propose a sufficient condition for fast convergence of NR in a nonlinear DFL network. This condition will be validated again in the diode clipper, along with providing an inequality for the input signal guaranteeing quadratic speed of convergence in this circuit. Finally, Sec. IV will extend the convergence analysis to the ring modulator by bringing to surface some practical limits, but also the potential of the proposed tool as a predictor of convergence speed in circuits containing multiple nonlinearities. Some remarks are made accordingly, in Sec. V, and conclusions are drawn in Sec. VI.

## II. LINEAR EQUIVALENT

When multiple scalar DFLs are present, the corresponding network can be computed by first aggregating all series and parallel DFL topologies, and then solving at every temporal step the  $N$ -dimensional equation  $\mathbf{v} = c(\mathbf{v}) + \mathbf{u}$ , where  $\mathbf{v}, \mathbf{u} \in \mathbb{R}^N$  and  $c : \mathbb{R}^N \rightarrow \mathbb{R}^N$ . If the network has memory meanwhile all nonlinear characteristics in the network are memoryless, then the output can be algebraically separated in two parts: the former is memoryless, in loop with the nonlinearities as explained in the beginning; the latter depends only on the memory of each linear block and hence participates in the previous equation as an additional component  $\mathbf{p}$ :  $\mathbf{v} = c(\mathbf{v}) + \mathbf{u} + \mathbf{p}$ . The state term  $\mathbf{p}$  can be consequently seen as an offset of the input  $\mathbf{u} = (\mathbf{u}_1, \dots, \mathbf{u}_N)^T$  and does not bring algebraic complications. For this reason, in the following we will not denote the state term explicitly whenever it can be aggregated in the input.

Furthermore, the scalar character of the network structure implies that all nonlinear blocks can be encapsulated in a function vector  $c = (c_1, \dots, c_N)^T$ , with  $c_i : \mathbb{R}^N \rightarrow \mathbb{R}$ . In fact,  $c_i$  contributes to the signal component  $\mathbf{v}_i$  by summing the outputs of the scalar nonlinear blocks  $c_{i,j} : \mathbb{R} \rightarrow \mathbb{R}$ , each processing the respective signal component  $\mathbf{v}_j$  [11]:

$$\mathbf{v}_i = c_i(\mathbf{v}) + \mathbf{u}_i = \sum_{j=1}^N c_{i,j}(\mathbf{v}_j) + \mathbf{u}_i. \quad (5)$$

A NR solution of the multiple DFL problem requires to search the zero of the  $N$ -dimensional function

$$f(\mathbf{v}) = \mathbf{v} - c(\mathbf{v}) - \mathbf{u} \quad (6)$$

by iterating on  $\mathbf{v}_{[k+1]} = \mathbf{v}_{[k]} - \mathbf{J}_f^{-1}(\mathbf{v}_{[k]})f(\mathbf{v}_{[k]})$ , in which  $\mathbf{J}_f = \left( \frac{\partial f_i}{\partial v_j} \right)$  is the Jacobian matrix of the function  $f$ . Such two formulas respectively generalize (2) and (1) to  $N$  dimensions. In this case the causal equivalent becomes too complicated, since any signal component in  $\mathbf{v}_{[k+1]}$  in general depends on all components in  $\mathbf{v}_{[k]}$  through the Jacobian matrix.

The following result formulates the causal equivalent at every NR iteration step  $k$  in terms of a delay-free *linear equivalent* network, which computes  $\mathbf{v}_{[k+1]}$  by solving a linear equation system in  $\mathbf{v}_{[k]}$  and  $\mathbf{u}$ . By means of this result, the original network is linearized in a structurally equivalent network that realizes the NR scheme.

At every step we first substitute each nonlinear block  $c_{i,j}$  in the original network with a linear approximation  $l_{i,j,[k]}$ , corresponding to the tangent to the respective function in  $\mathbf{v}_{j,[k]}$ :

$$l_{i,j,[k]}(\mathbf{v}_j) = c_{i,j}(\mathbf{v}_{j,[k]}) + c'_{i,j}(\mathbf{v}_{j,[k]})(\mathbf{v}_j - \mathbf{v}_{j,[k]}). \quad (7)$$

Then, we compute the array of signals  $\mathbf{v}_{[k+1]}$  by solving the linear system—see the scalar case in Fig. 1(c)

$$\mathbf{v} = l_{[k]}(\mathbf{v}) + \mathbf{u}, \quad (8)$$

in which  $l_{[k]}(\mathbf{v})$  is the array of block linearizations at NR iteration  $k$ :  $l_{[k]} = (l_{1,[k]}, \dots, l_{N,[k]})^T$ , with  $l_{i,[k]}(\mathbf{v}) = \sum_{j=1}^N l_{i,j,[k]}(\mathbf{v}_j)$ .

It can be easily proved that solving the linear DFL network expressed by (8) corresponds to compute a new NR iteration. By expanding (7) in (8) we obtain

$$\mathbf{v} = c(\mathbf{v}_{[k]}) + \mathbf{J}_c(\mathbf{v}_{[k]})(\mathbf{v} - \mathbf{v}_{[k]}) + \mathbf{u}. \quad (9)$$

The terms in (9) can be rearranged in the following formula:

$$\mathbf{v} = \mathbf{v}_{[k]} - (\mathbf{I} - \mathbf{J}_c(\mathbf{v}_{[k]}))^{-1}(\mathbf{v}_{[k]} - c(\mathbf{v}_{[k]}) - \mathbf{u}), \quad (10)$$

which implies  $\mathbf{v} = \mathbf{v}_{[k+1]}$ . In fact, from (6) it descends  $\mathbf{J}_f = \mathbf{I} - \mathbf{J}_c$ , hence (10) computes an NR iteration towards the zeros of the N-dimensional function  $f$ .

#### A. Case study: diode clipper

As a first example we solve the NR filter network representing a diode clipper. This circuit, shown in Fig. 2(a), contains a resistance  $R$  in series with a capacitance  $C$ , whose charge is alternatively leaked by two identical diodes in parallel to it, and in opposite orientation. Each diode in fact establishes a memoryless voltage-to-current nonlinearity  $g_D$ , admitting few or (in the limit) null current when the voltage is negative.

Equating the current coming from the resistor to the currents going to the capacitor and the diodes, we get—see Fig. 2(a)

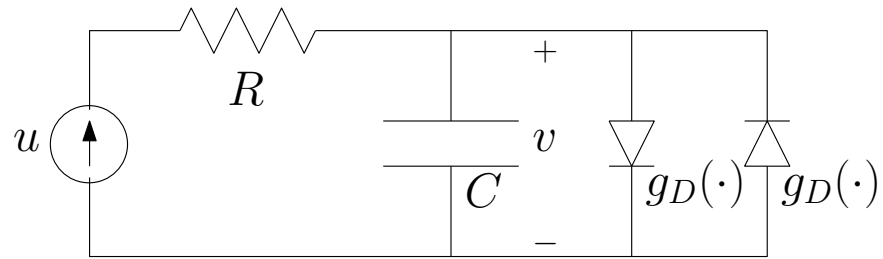
$$\frac{u - v}{R} = C \frac{dv}{dt} + g_D(v) - g_D(-v), \quad (11)$$

in which  $u$  is the input voltage and  $v$  is the voltage at the capacitor. Rearranging the terms and then integrating:

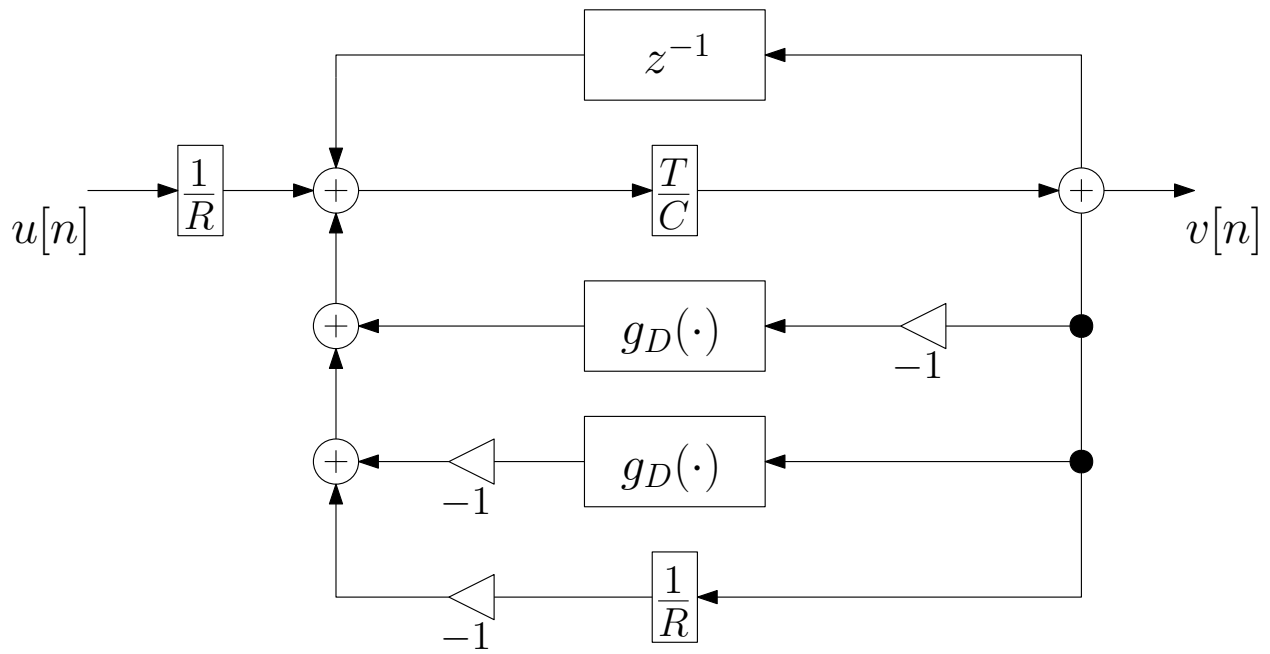
$$v = \frac{1}{C} \int \frac{1}{R} u - \frac{1}{R} v - g_D(v) + g_D(-v) dt. \quad (12)$$

Discretizing (12) through backward Euler with temporal step  $T$ , i.e.,  $y[n] = y[n-1] + Tx[n]$ , leads to the filter network in Fig. 2(b):

$$\begin{aligned} v[n] = \frac{T}{C} \left\{ \frac{u[n] - v[n]}{R} - g_D(v[n]) + g_D(-v[n]) \right\} \\ + v[n-1]. \end{aligned} \quad (13)$$



(a)



(b)

Fig. 2. Diode clipper: electronic circuit (a). Digital filter network using backward Euler discretization (b).

In this network the nonlinearities can be summed to form a single nonlinearity  $g_D(-v) - g_D(v)$ , reflecting the parallelism in the original circuit. Hence, the diode clipper gives rise to a scalar DFL network.

The linear equivalent is straightforwardly figured out for each  $v_{[k]}$ , by substituting  $g_D(v)$  with its linearization (7) in  $v_{[k]}$ :  $g_D(v_{[k]}) + g'_D(v_{[k]})(v - v_{[k]})$ . From here, at each temporal step  $n$  the

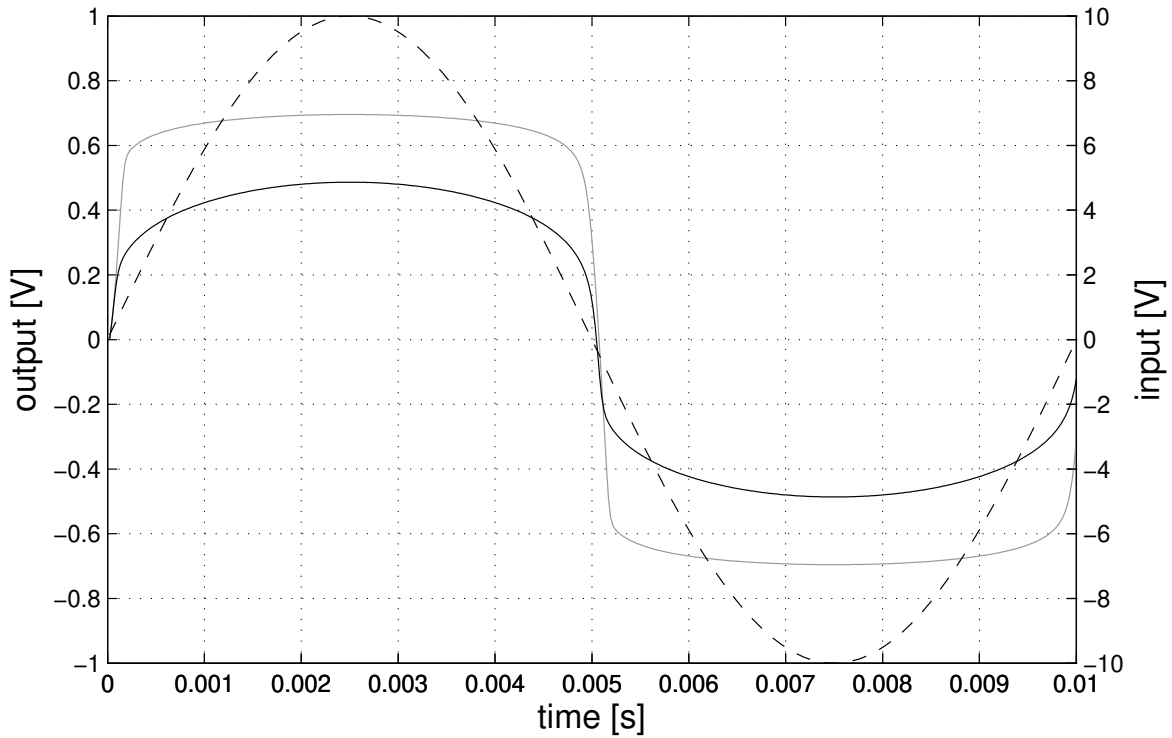


Fig. 3. Diode clipper. Responses to a 10 V sinusoid at 100 Hz (dashed line) using exponential (grey line) or polynomial (black line) diode characteristics.

$k$ th NR iteration is found which computes  $v_{[k+1]}$ :

$$v_{[k+1]} = \frac{v[n-1] + \frac{T}{RC}u[n] + \frac{T}{C}\{g'_D(v_{[k]}) + g'_D(-v_{[k]})\}v_{[k]}}{1 + \frac{T}{RC} + \frac{T}{C}g'_D(v_{[k]}) + \frac{T}{C}g'_D(-v_{[k]})} - \frac{\frac{T}{C}\{g_D(v_{[k]}) - g_D(-v_{[k]})\}}{1 + \frac{T}{RC} + \frac{T}{C}g'_D(v_{[k]}) + \frac{T}{C}g'_D(-v_{[k]})}$$

Fig. 3 (above) shows two responses of the diode clipper to a 10 V sinusoid oscillating at 100 Hz, shown in dashed line. The response in grey color results by using an exponential diode characteristic [32]  $i = g_D(v) = I_D(e^{v/2V_E} - 1)$  with  $I_D = 2.52$  nA and  $V_E = 23$  mV, whereas the response in black color results by using a polynomial characteristic [40]  $i = g_D(v) = V_P v^4 1(v)$ , in which  $1(v)$  is the unit step function in  $v = 0$  and  $V_P = 0.17$  A/V<sup>4</sup>. Furthermore,  $F_S = 1/T = 44100$  Hz,  $C = 100$  nF and  $R = 1$  k $\Omega$ . In both cases the iteration stops when  $|v_{[k+1]} - v_{[k]}| < 0.1$  mV.



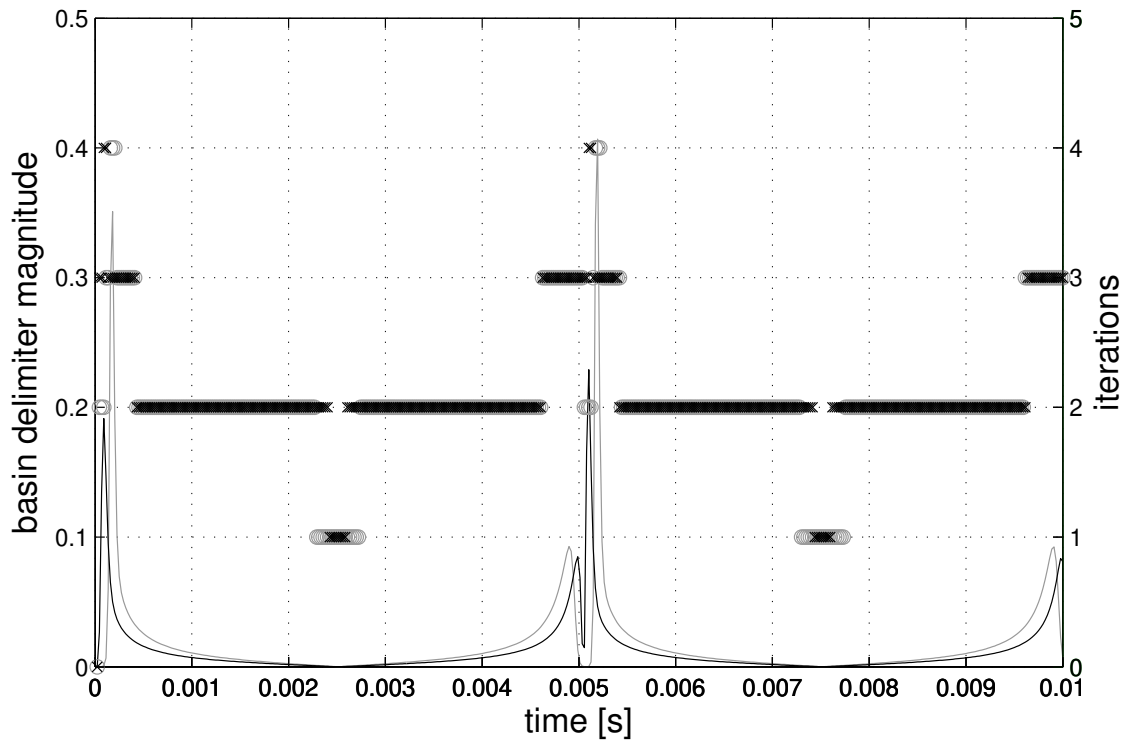


Fig. 4. Diode clipper. Number of NR iterations necessary to compute the responses of Fig. 3 when using exponential (grey circles) or polynomial (black crosses) diode characteristics. Corresponding basin delimiter in the exponential (grey curve) or polynomial (black curve) case.

### III. CONVERGENCE OF THE NETWORK

Fig. 4 shows corresponding numbers of iterations needed by NR to reach the stop condition, respectively using the exponential (grey circles) or polynomial (black crosses) characteristic. Predicting such numbers, or at least estimating their magnitude before the digital simulation of a nonlinear electronic audio device would be highly desirable. To this regard, a bound on NR iterations has been recently found for a class of collision models applicable to contact sounds synthesis [19]. The convergence speed of (also modified) Newton methods has been studied in Wave Digital Networks containing multiple nonlinearities, too [41]. Unfortunately, there is no general rule for counting in advance the iterations a NR solver will perform to find a root of (6). Even if some theoretical results have been obtained in the case when a function is polynomial [42], already cubic polynomials have been shown to have open sets of initial points whose boundaries are complicated fractals; NR does not lead to any root if starting from these

points, conversely it goes to an attracting cycle of period greater than one [43]. On the other hand, theorems exist providing sufficient conditions for convergence and fast convergence of the method [17]. The latter, in particular, define multidimensional *basins* inside which the solver converges quadratically to the solution  $\mathbf{v}$  in infinite norm:

$$M\|\mathbf{v} - \mathbf{v}_{[k]}\|_\infty \leq (M\|\mathbf{v} - \mathbf{v}_{[0]}\|_\infty)^{2^k}. \quad (14)$$

We will study these basins of fast convergence by adapting a known theorem [17] to our network models. In such models, in fact, (5) and (6) guarantee that each component  $f_i$  forming  $f(\mathbf{v})$  is a superposition of scalar functions  $f_{i,j}(\mathbf{v}_j)$ ,  $j = 1, \dots, N$ . Hence, the partial derivatives in the Jacobian  $\mathbf{J}_f$  reduce to ordinary derivatives and the same is true for the matrix  $\mathbf{H}_f$  of second derivatives of  $f(\mathbf{v})$ :

$$\mathbf{J}_f(\mathbf{v}) = (f'_{i,j}(\mathbf{v}_j)) \quad \text{and} \quad \mathbf{H}_f(\mathbf{v}) = (f''_{i,j}(\mathbf{v}_j)), \quad (15)$$

$i, j = 1, \dots, N$ .

Now, let us assume that  $\epsilon$  is such that  $f_{i,j}$ ,  $f'_{i,j}$  and  $f''_{i,j}$  are continuous in every interval  $\xi_j \in [\mathbf{v}_j - \epsilon, \mathbf{v}_j + \epsilon]$ ; furthermore, that  $\mathbf{J}_f$  has inverse in the hyper-rectangle  $I = \{\xi \in \mathbb{R}^N : \|\mathbf{v} - \xi\|_\infty < \epsilon\}$ . We form the set  $\mathcal{U}$  of matrices  $\mathbf{X} = (\xi_{i,j})$  whose rows belong to  $I$ , and define

$$M = \frac{1}{2} \max_{\xi \in I, \mathbf{X} \in \mathcal{U}} \|\mathbf{J}_f(\xi)^{-1} \mathbf{H}_f(\mathbf{X})\|_\infty, \quad (16)$$

where  $\mathbf{H}_f(\mathbf{X}) = (f''_{i,j}(\xi_{i,j}))$ . If  $\mathbf{v}_{[0]} \in I$  and

$$M\|\mathbf{v} - \mathbf{v}_{[0]}\|_\infty < 1, \quad (17)$$

then a NR iteration starting from  $\mathbf{v}_{[0]}$  generates a sequence that converges to  $\mathbf{v}$  with quadratic speed given by (14).

This result is proved in Appendix A. It is important to point out that the product  $M\|\mathbf{v} - \mathbf{v}_{[0]}\|_\infty$  cannot be computed prior to iterating the solver, since both its factors depend on the output  $\mathbf{v}$ . However, we will make convenient use of this tool by restricting the domain where  $M$  is evaluated to the trajectory  $\mathbf{v}[n]$  of the solution:

$$M(\mathbf{v}) = \frac{1}{2} \|\mathbf{J}_f(\mathbf{v})^{-1} \mathbf{H}_f(\mathbf{v})\|_\infty, \quad (18)$$

in which the maximum search across  $\xi \in I$ ,  $\mathbf{X} \in \mathcal{U}$  appearing in (16) has been removed. The consequently modified product  $S(\mathbf{v}) = M(\mathbf{v})\|\mathbf{v} - \mathbf{v}_{[0]}\|_\infty$  will be called *basin delimiter* from here on. We will use it by assuming  $\mathbf{v}_{[0]} = \mathbf{v}[n-1]$ , since the NR solution at step  $n$  is typically (but not necessarily [41]) searched by starting the iteration at the previous solution point.

### A. Case study: diode clipper

For instance in the diode clipper, Eq. (6) is in particular equal to (2):

$$\begin{aligned} f(v) &= v - c(v) - u \\ &= v - \rho(g_D(-v) - g_D(v) - F_S C v[n-1] + \frac{1}{R}u[n]). \end{aligned} \quad (19)$$

with  $\rho = R/(1 + RCF_S)$ . From here  $S(v)$  is immediately figured out by computing (18) in the scalar case:

$$M(v) = \frac{1}{2} \left| \frac{f''(v)}{f'(v)} \right| = \frac{1}{2} \left| \frac{\rho(g_D''(v) - g_D''(-v))}{1 + \rho(g_D'(v) + g_D'(-v))} \right|. \quad (20)$$

Fig. 4 shows the basin delimiter for the diode clipper simulation seen in Sec. II-A. As hypothesized, it follows the number of NR iterations with good accuracy across simulation time. Moreover its magnitude is always smaller than one, suggesting quadratic convergence everywhere including more critical regions, e.g. when the steepness of the output signal changes faster.

A closer look to (20) reveals that  $M(v)$  is limited. In fact, when  $v > 0$  the part of  $f(v)$  which is free from the constant offset  $u$  depending on the input and state, namely

$$\begin{aligned} f_u(v) &= f(v) + u = v - c(v) \\ &= v + \rho(g_D(v) - g_D(-v)), \end{aligned} \quad (21)$$

is accurately approximated by  $\tilde{f}_u(v) = v + \rho g_D(v)$ , which is still always positive in this voltage range. Hence, continuing with this approximation,  $f'(v) = \tilde{f}_u'(v)$  and finally

$$M'(v) = \frac{1}{2} \frac{\tilde{f}_u'''(v)\tilde{f}_u'(v) - (\tilde{f}_u''(v))^2}{(\tilde{f}_u'(v))^2}, \quad v > 0. \quad (22)$$

From here, a look at where  $\tilde{f}_u'''(v)\tilde{f}_u'(v) = (\tilde{f}_u''(v))^2$  shows that  $M(v)$  has two symmetric maxima equal to

$$\begin{aligned} M(\infty) &= \frac{1}{4V_E} \quad \text{and} \\ M\left(\sqrt[3]{\frac{1}{2\rho V_P}}\right) &= \sqrt[3]{2\rho V_P}, \end{aligned} \quad (23)$$

respectively if the diode characteristic is exponential or polynomial.

Limitation of  $M(v)$  is especially desirable under the hypothesis of *global* convergence of the scheme to a bounded-energy solution  $v$ . This hypothesis implies that  $|v - v_{[0]}|$  is in its turn limited. Hence, quadratic speed of convergence can be guaranteed by increasing  $F_S$  until

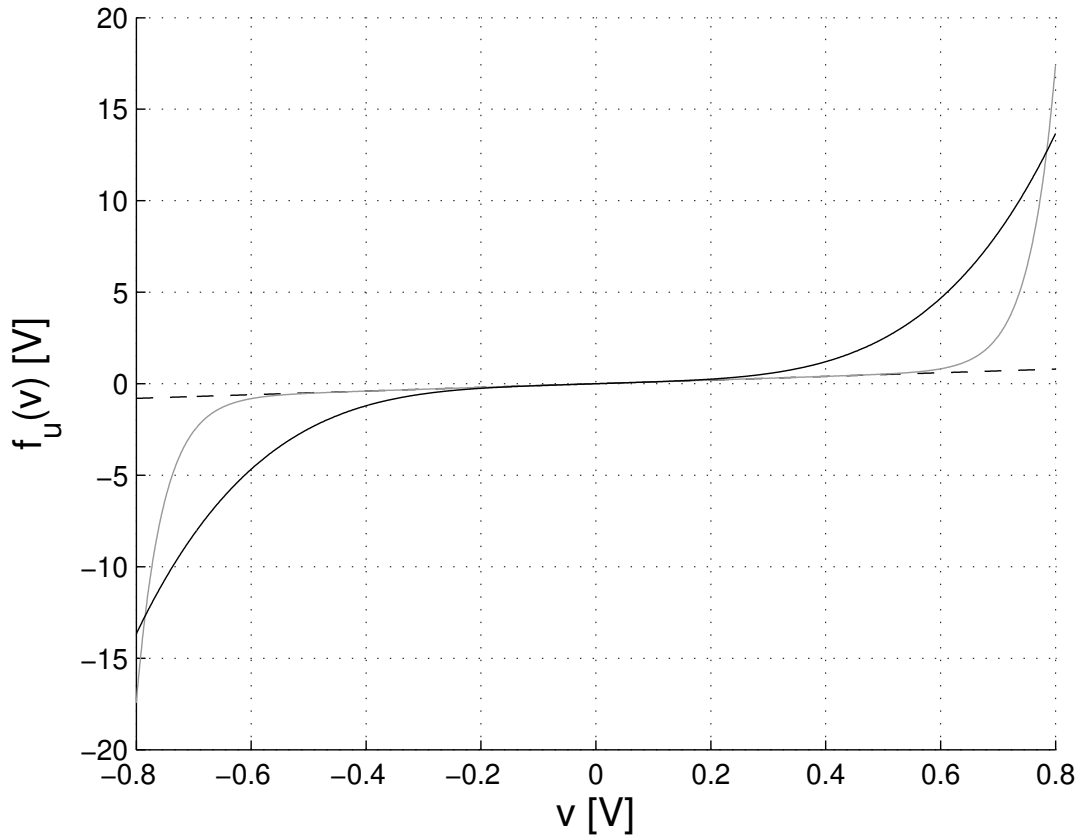


Fig. 5. Diode clipper. Global NR convergence with exponential (grey curve) or polynomial (black curve) diode characteristics. Diagonal function in dashed line given for reference.

$S(v)$  satisfies inequality (17). In fact, proportionally downscaling the temporal step reduces the distance  $|v - v_{[0]}|$  and, by (23) in the case of polynomial diode characteristic, also  $M(v)$ . Now, the energy in the diode clipper is finite [32], furthermore global convergence of the NR solution is guaranteed [43] by checking that  $f_u(v)$  lies in absolute terms above the diagonal and furthermore has increasing derivative [41]. This fact is shown in Fig. 5. In conclusion, there exist a value  $F_S$  beyond which the NR solver convergences globally with quadratic speed.

Exceptionally in the case of the diode clipper we have been able to derive an inequality that limits the output within known values of the input:

$$\begin{aligned}
 |v[n+1] - v[n]| &\leq \max_{1 < k < n} |u[k+1] - u[k]| \\
 &+ (\rho C F_S)^n \left\{ |v[1] - v[0]| - \max_{1 < k < n} |u[k+1] - u[k]| \right\}.
 \end{aligned} \tag{24}$$

The derivation is shown in Appendix B. By guaranteeing that after an initial transient, which

is represented by the exponentially decaying term in (24), subsequent output values are always closer than the largest subsequent pair of input values so far, in practice this inequality can be used to bound the output and, hence, to ensure quadratic convergence.

All the results discussed so far generalize to polynomial functions  $g(v) = \sum_{i=1}^P K_i v^i 1(v)$  in which  $K_i \geq 0$ . More in general, the assumption  $g'(v) \geq 0$  implies  $f'(v) \geq 1$ . Under this condition NR is at least locally convergent and (24) continues to hold. On the other hand it is simple to note that the function  $g(v) = e^{v^2} 1(v)$  is such that  $M(v)$  is not bounded anymore. Once again it is recalled that (17) provides a sufficient condition of fast convergence: failing to hit it does not imply that the NR solution will converge slowly.

Since limited to the scalar case, this result is valid for circuits containing  $P$  diodes in parallel. In particular it cannot be extended to diode- or transistor-based ladder circuits such as the Moog or VCS3 voltage-controlled filter, whose equivalent DFL network contains hyperbolic tangent characteristics [10], [15]. For such characteristics in fact the resulting function  $f_u(v)$  does not lie above the diagonal as in Fig. 5, and for this reason convergence of the NR solution cannot be guaranteed. On the other hand the same result completes the proof of convergence that has been given in a recent work on Wave Digital Networks, whose nonlinear characteristics were systematically computed one by one through NR by splitting the multidimensional nonlinearity (6) in a set of scalar functions (2) [38].

In the same work the Wave Digital model was successfully tested on a ring modulator circuit [40]. Similarly to the diode clipper, the ring modulator has been often considered in the literature of virtual analog—some examples of its use are listed in [38, Sec. 1]. With slower performance, this circuit had been previously computed also by implementing a fixed-point method on its equivalent DFL network, as part of a general proof of convergence of this solver on such networks [11]. It seems logical, at this point, to continue our investigation on the ring modulator in an aim to generalize the properties we have found for the diode clipper.

#### IV. CONVERGENCE IN THE RING MODULATOR

Thanks to the closed-loop connection of four diodes, the ring modulator analogue model shown in Fig. 6 generates an output  $v_2(t)$  by multiplying (in an analogue sense) two inputs  $m(t)$  and  $c(t)$ , respectively a modulator and a carrier voltage signal. This model was obtained by loading the output point of the original circuit with a resistance  $R_a$ , and then by putting a carrier source resistance  $R_i$  in parallel with a regularizing capacitance  $C_p$  [40]. As a result, it

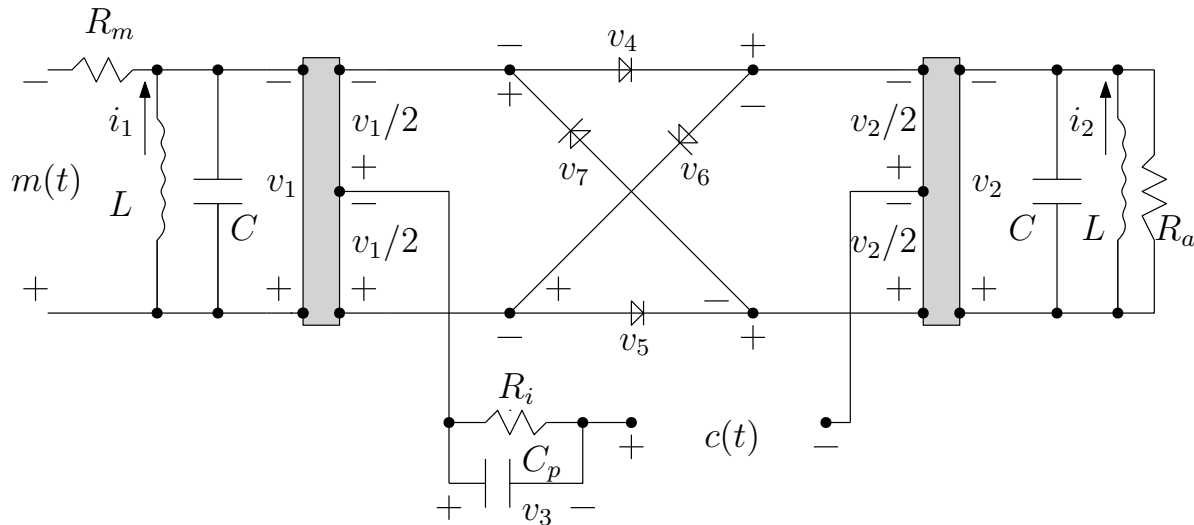


Fig. 6. Ring modulator analogue model [40].

leads to a system of two current and seven voltage ordinary differential equations in which the voltages  $v_1$ ,  $v_2$  at the transformers and  $v_3$  in series with the carrier signal are instantaneously fed back to the equations computing the currents  $i_1$ ,  $i_2$  at the transformers and the voltages  $v_4$ ,  $v_5$ ,  $v_6$ ,  $v_7$  at the diodes. Turning such equations to the discrete-time domain using backward Euler at temporal step  $n$  leads to the following implicit system of nine difference equations in nine

unknown variables:

$$\begin{aligned}
v_1 &= \rho_m \left( \frac{m}{R_m} + i_1 - \frac{g(v_4)}{2} + \frac{g(v_5)}{2} - \frac{g(v_6)}{2} + \frac{g(v_7)}{2} \right) \\
&\quad + \rho_m C F_s v_1 [n - 1] \\
v_2 &= \rho_a \left( i_2 + \frac{g(v_4)}{2} - \frac{g(v_5)}{2} - \frac{g(v_6)}{2} + \frac{g(v_7)}{2} \right) \\
&\quad + \rho_a C F_s v_2 [n - 1] \\
v_3 &= \rho_i (g(v_4) + g(v_5) - g(v_6) - g(v_7)) \\
&\quad + \rho_i C_p F_s v_3 [n - 1] \\
v_4 &= \frac{v_1}{2} - \frac{v_2}{2} - v_3 - c \\
v_5 &= -\frac{v_1}{2} + \frac{v_2}{2} - v_3 - c \\
v_6 &= \frac{v_1}{2} + \frac{v_2}{2} + v_3 + c \\
v_7 &= -\frac{v_1}{2} - \frac{v_2}{2} + v_3 + c \\
i_1 &= -\frac{1}{L F_s} v_1 + i_1 [n - 1] \\
i_2 &= -\frac{1}{L F_s} v_2 + i_2 [n - 1]
\end{aligned} \tag{25}$$

in which

$$\rho_m = \frac{R_m}{1 + R_m C F_s}, \quad \rho_a = \frac{R_a}{1 + R_a C F_s}, \quad \rho_i = \frac{R_i}{1 + R_i C_p F_s}.$$

Perhaps interestingly, the ring modulator is equivalent to the diode clipper of Sec. II-A in presence of constant inputs. In fact, dc voltages short-circuit all inductors and open the capacitors in ways that  $m$  is isolated within a mesh containing only the resistance  $R_m$ , and  $c$  flows along resistance  $R_i$  and, in series with it, diodes  $D_4$  and  $D_5$  oriented in one direction as well as diodes  $D_6$  and  $D_7$  oriented in the opposite direction. Hence, the voltage  $v$  across the diodes obeys to the same equation as (11), with null capacitances and double diode currents:

$$v + 2R_i (g_D(v) - g_D(-v)) - c = 0. \tag{26}$$

Basin delimiters needs to calculate  $M(\mathbf{v})$ . Recalling (18) and holding (25), the Jacobian  $\mathbf{J}_f(\mathbf{v}) = \mathbf{I} - \mathbf{J}_c(\mathbf{v})$  turns out to be equal to (27). Its block-based structure of the type  $\mathbf{J}_f =$

$$\mathbf{J}_f(\mathbf{v}) = \begin{bmatrix} 1 & 0 & 0 & \frac{\rho_m}{2} g'_D(v_4) & -\frac{\rho_m}{2} g'_D(v_5) & \frac{\rho_m}{2} g'_D(v_6) & -\frac{\rho_m}{2} g'_D(v_7) & -\rho_m & 0 \\ 0 & 1 & 0 & -\frac{\rho_a}{2} g'_D(v_4) & \frac{\rho_a}{2} g'_D(v_5) & \frac{\rho_a}{2} g'_D(v_6) & -\frac{\rho_a}{2} g'_D(v_7) & 0 & -\rho_a \\ 0 & 0 & 1 & -\rho_l g'_D(v_4) & -\rho_l g'_D(v_5) & \rho_l g'_D(v_6) & \rho_l g'_D(v_7) & 0 & 0 \\ -\frac{1}{2} & \frac{1}{2} & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & -\frac{1}{2} & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ -\frac{1}{2} & -\frac{1}{2} & -1 & 0 & 0 & 1 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & -1 & 0 & 0 & 0 & 1 & 0 & 0 \\ \frac{1}{LF_s} & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & \frac{1}{LF_s} & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (27)$$

$$\mathbf{A}(\mathbf{v}) = \begin{bmatrix} -\frac{\rho_m}{2} g'_D(v_4) & \frac{\rho_m}{2} g'_D(v_5) & -\frac{\rho_m}{2} g'_D(v_6) & \frac{\rho_m}{2} g'_D(v_7) & \rho_m & 0 \\ \frac{\rho_a}{2} g'_D(v_4) & -\frac{\rho_a}{2} g'_D(v_5) & -\frac{\rho_a}{2} g'_D(v_6) & \frac{\rho_a}{2} g'_D(v_7) & 0 & \rho_a \\ \rho_l g'_D(v_4) & \rho_l g'_D(v_5) & -\rho_l g'_D(v_6) & -\rho_l g'_D(v_7) & 0 & 0 \end{bmatrix} \quad (28)$$

$\begin{bmatrix} \mathbf{I} & -\mathbf{A} \\ -\mathbf{B} & \mathbf{I} \end{bmatrix}$  allows for the following decomposition:

$$\mathbf{J}_f = \begin{bmatrix} \mathbf{I} & \mathbf{O} \\ -\mathbf{B} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{O} \\ \mathbf{O} & \mathbf{I} - \mathbf{BA} \end{bmatrix} \begin{bmatrix} \mathbf{I} & -\mathbf{A} \\ \mathbf{O} & \mathbf{I} \end{bmatrix}, \quad (29)$$



with  $\mathbf{A}$  as in (28), and

$$\mathbf{B} = \begin{bmatrix} 1/2 & -1/2 & -1 \\ -1/2 & 1/2 & -1 \\ 1/2 & 1/2 & 1 \\ -1/2 & -1/2 & 1 \\ -\frac{1}{LF_s} & 0 & 0 \\ 0 & -\frac{1}{LF_s} & 0 \end{bmatrix}. \quad (30)$$

Hence,  $\mathbf{J}_f$  has an inverse if and only if  $\mathbf{I} - \mathbf{BA} \in \mathbb{R}^{3 \times 3}$  has an inverse. Moreover, the inversion of  $\mathbf{J}_f$  is reduced to the problem of calculating  $(\mathbf{I} - \mathbf{BA})^{-1}$ :

$$\mathbf{J}_f^{-1} = \begin{bmatrix} \mathbf{I} + \mathbf{A}(\mathbf{I} - \mathbf{BA})^{-1}\mathbf{B} & -\mathbf{A}(\mathbf{I} - \mathbf{BA})^{-1} \\ -(\mathbf{I} - \mathbf{BA})^{-1}\mathbf{B} & (\mathbf{I} - \mathbf{BA})^{-1} \end{bmatrix}. \quad (31)$$

The Sherman-Morrison-Woodbury formula [44] yields

$$(\mathbf{I} - \mathbf{BA})^{-1} = \mathbf{I} + \mathbf{B}(\mathbf{I} - \mathbf{AB})^{-1}\mathbf{A}.$$

It follows that

$$\begin{aligned} (\mathbf{I} - \mathbf{BA})^{-1}\mathbf{B} &= \mathbf{B} + \mathbf{B}(\mathbf{I} - \mathbf{AB})^{-1}\mathbf{AB} \\ &= \mathbf{B}(\mathbf{I} + (\mathbf{I} - \mathbf{AB})^{-1}\mathbf{AB}) = \mathbf{B}(\mathbf{I} - \mathbf{AB})^{-1} \end{aligned}$$

and, considering also to swap  $\mathbf{A}$  and  $\mathbf{B}$  in such two equations,

$$\mathbf{J}_f^{-1} = \begin{bmatrix} (\mathbf{I} - \mathbf{AB})^{-1} & (\mathbf{I} - \mathbf{AB})^{-1}\mathbf{A} \\ \mathbf{B}(\mathbf{I} - \mathbf{AB})^{-1} & \mathbf{I} + \mathbf{B}(\mathbf{I} - \mathbf{AB})^{-1}\mathbf{A} \end{bmatrix}. \quad (32)$$

If we define

$$\mathbf{D}_\rho = \text{diag}(\rho_m, \rho_a, \rho_l)$$

$$\mathbf{D}_{g'_D} = \text{diag}(g'_D(v_4), g'_D(v_5), g'_D(v_6), g'_D(v_7))$$

$$\mathbf{W} = \begin{bmatrix} -1/2 & 1/2 & 1 \\ 1/2 & -1/2 & 1 \\ -1/2 & -1/2 & -1 \\ 1/2 & 1/2 & -1 \end{bmatrix}, \quad (33)$$

$$\mathbf{E} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \quad \text{such that } \mathbf{E}\mathbf{E}^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

then

$$\begin{aligned} \mathbf{A} &= \mathbf{D}_\rho \begin{bmatrix} \mathbf{W}^T \mathbf{D}_{g'_D} & \mathbf{E} \end{bmatrix} \\ \mathbf{B} &= - \begin{bmatrix} \mathbf{W} \\ \frac{1}{LF_s} \mathbf{E}^T \end{bmatrix}. \end{aligned} \quad (34)$$

Hence,

$$\mathbf{I} - \mathbf{AB} = \mathbf{I} + \mathbf{D}_\rho (\mathbf{W}^T \mathbf{D}_{g'_D} \mathbf{W} + \frac{1}{LF_s} \mathbf{E} \mathbf{E}^T)$$

contains the symmetric positive semidefinite matrix  $\mathbf{W}^T \mathbf{D}_{g'_D} \mathbf{W} + \mathbf{E} \mathbf{E}^T / (LF_s)$ , for  $\mathbf{D}_{g'_D}$  has always nonnegative diagonal entries. This implies that its eigenvalues are real and nonnegative. Now, the similar matrix

$$\begin{aligned} \mathbf{D}_\rho^{-1/2} (\mathbf{I} - \mathbf{AB}) \mathbf{D}_\rho^{1/2} &= \\ \mathbf{I} + \mathbf{D}_\rho^{1/2} (\mathbf{W}^T \mathbf{D}_{g'_D} \mathbf{W} + \frac{1}{LF_s} \mathbf{E} \mathbf{E}^T) \mathbf{D}_\rho^{1/2} \end{aligned} \quad (35)$$

has obviously the same eigenvalues as  $\mathbf{I} - \mathbf{AB}$ , furthermore it is the sum of an identity matrix plus a matrix that is in its turn symmetric positive semidefinite. From the similarity (35) it can be concluded that  $\mathbf{I} - \mathbf{AB}$  has eigenvalues that are greater than or equal to one. As an immediate consequence such a matrix has always an inverse.

This conclusion extends the property we showed in Sec. III-A for the diode clipper, where  $g'(v) \geq 0$  implied  $f'(v) \geq 1$  and, hence, computability of the NR scheme.

Concerning inversion, from (35) it descends

$$\begin{aligned} (\mathbf{I} - \mathbf{AB})^{-1} &= \mathbf{D}_\rho^{1/2} (\mathbf{D}_\rho^{-1/2} (\mathbf{I} - \mathbf{AB}) \mathbf{D}_\rho^{1/2})^{-1} \mathbf{D}_\rho^{-1/2} = \\ \mathbf{D}_\rho^{1/2} (\mathbf{I} + \mathbf{D}_\rho^{1/2} (\mathbf{W}^T \mathbf{D}_{g'_D} \mathbf{W} + \frac{1}{LF_s} \mathbf{E} \mathbf{E}^T) \mathbf{D}_\rho^{1/2})^{-1} \mathbf{D}_\rho^{-1/2}. \end{aligned}$$

The norm of the inverse, hence, can be split in three factors with the norm of (35) in the middle. Since the eigenvalues are smaller than or equal to one, the 2-norm of (35) is smaller than or equal to one in its turn. In fact, for symmetric matrices this norm is equal to the spectral radius [45]. If the mid factor is removed from the norm of the inverse, then

$$\begin{aligned} \|(\mathbf{I} - \mathbf{AB})^{-1}\|_2 &\leq \|\mathbf{D}_\rho^{1/2}\|_2 \|\mathbf{D}_\rho^{-1/2}\|_2 = \\ &= \sqrt{\|\mathbf{D}_\rho\|_2} \sqrt{\|\mathbf{D}_\rho^{-1}\|_2} = \sqrt{\frac{\max\{\rho_m, \rho_a, \rho_l\}}{\min\{\rho_m, \rho_a, \rho_l\}}}. \end{aligned} \quad (36)$$

$$\mathbf{C}(\mathbf{v}) = \begin{bmatrix} -\frac{\rho_m}{2}g_D''(v_4) & \frac{\rho_m}{2}g_D''(v_5) & -\frac{\rho_m}{2}g_D''(v_6) & \frac{\rho_m}{2}g_D''(v_7) & 0 & 0 \\ \frac{\rho_a}{2}g_D''(v_4) & -\frac{\rho_a}{2}g_D''(v_5) & -\frac{\rho_a}{2}g_D''(v_6) & \frac{\rho_a}{2}g_D''(v_7) & 0 & 0 \\ \rho_l g_D''(v_4) & \rho_l g_D''(v_5) & -\rho_l g_D''(v_6) & -\rho_l g_D''(v_7) & 0 & 0 \end{bmatrix} \quad (38)$$

Furthermore, if  $\mathbf{M} \in \mathbb{R}^{N \times N}$  then inequality  $\|\mathbf{M}\|_\infty \leq \sqrt{N} \|\mathbf{M}\|_2$  holds for the  $\infty$ -norm, which is equal to the largest sum chosen among the entries' absolute values forming each row [46]:  $\|\mathbf{M}\|_\infty = \max_i \sum_j |M_{i,j}|$ . From here,

$$\|(\mathbf{I} - \mathbf{AB})^{-1}\|_\infty \leq \sqrt{3 \frac{\max\{\rho_m, \rho_a, \rho_l\}}{\min\{\rho_m, \rho_a, \rho_l\}}}. \quad (37)$$

In order to formulate the basin delimiter for the ring modulator we also need to compute  $\mathbf{H}_f(\mathbf{v})$ , which has the following block structure:

$$\mathbf{H}_f = \begin{bmatrix} \mathbf{O} & \mathbf{C} \\ \mathbf{O} & \mathbf{O} \end{bmatrix}, \quad (39)$$

with  $\mathbf{O}$  a null matrix and  $\mathbf{C}$  as in (38). Hence, immediately from (31),

$$\mathbf{J}_f(\mathbf{v})^{-1} \mathbf{H}_f(\mathbf{v}) = \begin{bmatrix} \mathbf{O} & (\mathbf{I} - \mathbf{AB})^{-1} \mathbf{C} \\ \mathbf{O} & -\mathbf{B}(\mathbf{I} - \mathbf{AB})^{-1} \mathbf{C} \end{bmatrix}. \quad (40)$$

Since  $\|\mathbf{B}\|_\infty \leq \max\{2, 1/(LF_S)\}$ , then  $M(\mathbf{v})$  is figured out by the lower row in (40):

$$\|\mathbf{J}_f(\mathbf{v})^{-1} \mathbf{H}_f(\mathbf{v})\|_\infty = \|\mathbf{B}\|_\infty \|(\mathbf{I} - \mathbf{AB})^{-1}\|_\infty \|\mathbf{C}\|_\infty \quad (41)$$

Unfortunately though,  $\|\mathbf{C}\|_\infty$  depends on  $g_D''(v)$ . In fact, from (38):

$$\begin{aligned} \|\mathbf{C}\|_\infty &\leq 4 \max\{\rho_m, \rho_a, \rho_l\} \max_{i=4,5,6,7} \{g_D''(v_i)\} \\ &= 4 \max\{\rho_m, \rho_a, \rho_l\} g_D''(\max_{i=4,5,6,7} \{v_i\}), \end{aligned} \quad (42)$$

in which the latter equality descends from the monotonicity of  $g''$ . Contrarily to the diode clipper, for which  $M(v)$  was shown to be bounded by (23) for both the exponential and polynomial characteristics, in the case of the ring modulator inequality (42) in principle does not set a limit for  $M(\mathbf{v})$ . Nor efforts aimed at maintaining the magnitude of  $g_D'$  inside  $\|(\mathbf{I} - \mathbf{AB})^{-1}\|_\infty$  so far

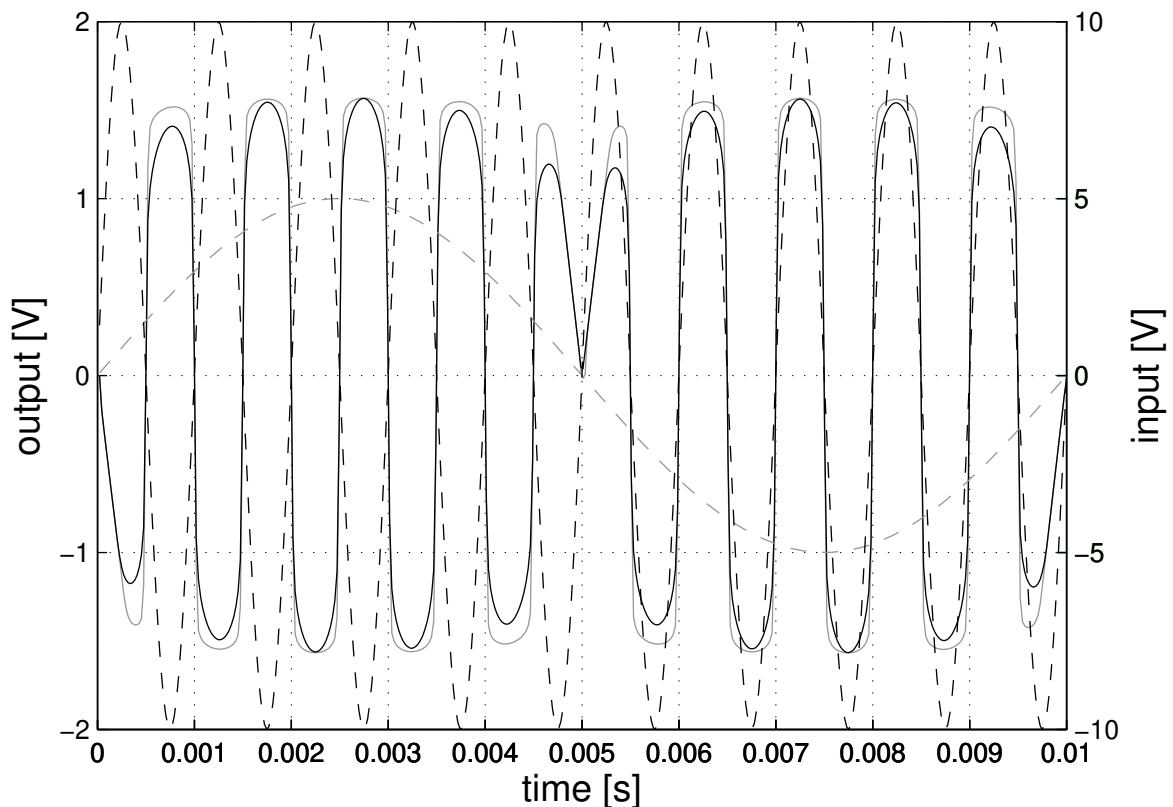


Fig. 7. Ring modulator. Responses to a 10 V modulating sinusoid at 1000 Hz (dashed black line) and 5 V carrier sinusoid at 100 Hz (dashed grey line) using exponential (grey solid line) or polynomial (black solid line) diode characteristics.

led us to a formulation of the basin delimiter preserving, in the case of the ring modulator, the counterbalancing role such first derivatives instead had in the scalar case—see Eq. (20).

The simulations in the next section will show that (41), although probably overestimating  $M(\mathbf{v})$  in the case of the ring modulator, nevertheless warns about potential drifts from quadratic convergence that, in practice, manifest if the exponential diode characteristic is chosen.

#### A. Simulations

Fig. 7 shows the responses to a 10 V modulating sinusoid at 1000 Hz, in presence of a 5 V carrier oscillating at 100 Hz with  $C = C_p = 10^{-9}$  F,  $L = 0.8$  H,  $R_a = 600$   $\Omega$ ,  $R_i = 50$   $\Omega$ ,  $R_m = 80$   $\Omega$ . The simulation runs at 44.1 kHz.

Fig. 8 shows that with the above parameters the basin delimiter  $S(\mathbf{v})$  is always greater than one, using either diode characteristic. Though, in the exponential case  $S(\mathbf{v})$  is clearly larger—

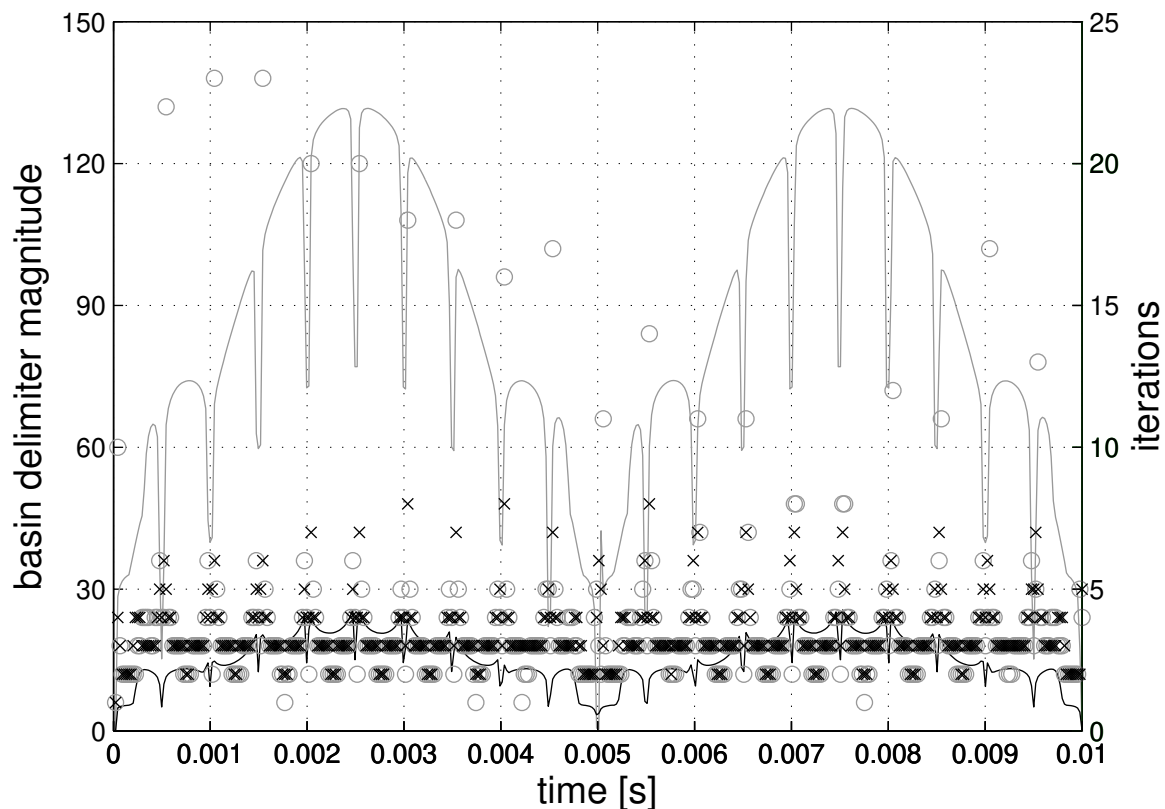


Fig. 8. Ring modulator. Number of NR iterations necessary to compute the responses of Fig. 7 when using exponential (grey circles) or polynomial (black crosses) diode characteristics. Corresponding basin delimiter in the exponential (grey curve) or polynomial (black curve) case.

compare the functions in grey and black line. This difference suggests that the NR solver may converge more slowly in this case, possibly with occasional vacancies from quadratic speed. Such vacancies do happen in correspondence of temporal steps whose iterations, otherwise normally less than ten, suddenly jump up to fifteen and more.

A deeper look to Fig. 8 shows that these drifts occur immediately after characteristic notches, affecting an otherwise oscillatory evolution of the basin delimiter. This evidence may bring interesting implications to such systems design when the real-time constraint holds, as abrupt drops from a regular trajectory of the basin delimiter may signal possible slowdown of the NR solver at the next temporal step. Possible explanations of this evidence are left to future research.

## V. FINAL REMARKS

DFL networks provide direct access to the (lumped) variables that circulate in the nonlinear model. For what we have seen in Sec. II, they admit a straightforward NR solution of the circuit which can be easily turned in an automatic computer procedure, for instance by organizing the linearly equivalent DFL topology and its blocks in proper matrix representations [47]. Another advantage they offer is the independence of the discretization method: as far as the nonlinearities are memoryless, the analog-to-digital transformation affects only the memory of the linear blocks, with minimal or no changes in the NR solver.

On the other hand, the lumped-variable approach is known to suffer from energy issues even if the network contains only passive elements. These issues can cause numerical instability as opposed to Wave Digital Networks, which are passive-guaranteed instead [4]. The proposed DFL networks are not exempt from energy issues, however it should be noted that they are general enough to model also passive-guaranteed networks.

An inherent limitation of DFL networks consists of the dimension of each nonlinear characteristics, which must be scalar as any other block in the network. Even a geometric nonlinearity  $c(\boldsymbol{v}) = v_1 v_2$ , which is two-dimensional, cannot find place on them and must be substituted by a composition of polynomials:  $2v_1 v_2 = (v_1 + v_2)^2 - v_1^2 - v_2^2$ . Kolmogorov's superposition theorem allows for substituting any multivariate function with a sum and composition of monovariate functions [48]. Clearly, substitutions of this kind are not computationally convenient.

## VI. CONCLUSIONS

We have carried on research about the fast computability of nonlinear DFL networks, by investigating properties of the NR scheme which for its quadratic speed of convergence is frequently used in the simulation of electronic circuits containing nonlinear characteristics. Specifically, we have first shown that NR can be directly applied to the network through a linearization of its nonlinear blocks. Then, we have found sufficient conditions for quadratic convergence in such networks depending on the magnitude of the basin delimiter, a distance function we have derived from a known theorem of scalar NR convergence. Even though it cannot be readily employed as a predictive tool, in this paper the basin delimiter has been used to figure out conditions guaranteeing quadratic convergence in the diode clipper. Furthermore, its application to the ring modulator has confirmed proportionality of its magnitude with the speed

of NR convergence, and possible predictive behavior depending on magnitude discontinuities which will be object of future research.

#### ACKNOWLEDGMENT

The authors acknowledge the support of the PRID project ENCASE funded by the University of Udine.

#### APPENDIX A

##### PROOF OF QUADRATIC CONVERGENCE (14)

If  $\mathbf{v}_{[0]} = (\mathbf{v}_{1,[0]}, \dots, \mathbf{v}_{N,[0]}) \in I$  then we can write the Taylor series of each function component  $f_{i,j}$  around the solution up to the quadratic term:

$$\begin{aligned} f_{i,j}(\mathbf{v}_j) &= f_{i,j}(\mathbf{v}_{j,[0]}) + f'_{i,j}(\mathbf{v}_{j,[0]})(\mathbf{v}_j - \mathbf{v}_{j,[0]}) \\ &\quad + \frac{1}{2}f''_{i,j}(\boldsymbol{\xi}_{i,j,[0]})(\mathbf{v}_j - \mathbf{v}_{j,[0]})^2, \end{aligned} \quad (43)$$

where  $\boldsymbol{\xi}_{i,j,[0]}$  lies between  $\mathbf{v}_j$  and  $\mathbf{v}_{j,[0]}$ . By summing over  $j$  we obtain

$$\begin{aligned} \sum_{j=1}^N f_{i,j}(\mathbf{v}_j) &= \sum_{j=1}^N f_{i,j}(\mathbf{v}_{j,[0]}) + \sum_{j=1}^N f'_{i,j}(\mathbf{v}_{j,[0]})(\mathbf{v}_j - \mathbf{v}_{j,[0]}) \\ &\quad + \frac{1}{2} \sum_{j=1}^N f''_{i,j}(\boldsymbol{\xi}_{i,j,[0]})(\mathbf{v}_j - \mathbf{v}_{j,[0]})^2. \end{aligned}$$

Since  $\sum_{j=1}^N f_{i,j}(\mathbf{v}_j) = 0$  for all  $i$ , then

$$f(\mathbf{v}_{[0]}) + \mathbf{J}_f(\mathbf{v}_{[0]})(\mathbf{v} - \mathbf{v}_{[0]}) + \frac{1}{2}\mathbf{H}_f(\boldsymbol{\xi}_{[0]})(\mathbf{v} - \mathbf{v}_{[0]})^2 = \mathbf{0},$$

in which  $\mathbf{0}$  is a null vector and  $(\mathbf{v} - \mathbf{v}_{[0]})^2$  contains the squares of each entry forming  $\mathbf{v} - \mathbf{v}_{[0]}$ .

Recalling (10), the previous formula can be rewritten as

$$\begin{aligned} \|\mathbf{v} - \mathbf{v}_{[1]}\|_\infty &\leq \frac{1}{2} \|\mathbf{J}_f(\mathbf{v}_{[0]})^{-1} \mathbf{H}_f(\boldsymbol{\xi}_{[0]})\|_\infty \|\mathbf{v} - \mathbf{v}_{[0]}\|_\infty^2 \\ &\leq M \|\mathbf{v} - \mathbf{v}_{[0]}\|_\infty^2, \end{aligned}$$

hence  $M \|\mathbf{v} - \mathbf{v}_{[1]}\|_\infty \leq (M \|\mathbf{v} - \mathbf{v}_{[0]}\|_\infty)^2 < 1$ . From here, by induction we obtain that all following iterates lie within  $I$ , furthermore (14) is straightforwardly derived. This guarantees quadratic convergence to the solution  $\mathbf{v}$  of the initial iteration starting in  $\mathbf{v}_{[0]}$  under the hypotheses given in Sec. III.

## APPENDIX B

## DIODE CLIPPER: INPUT CONSTRAINT FOR QUADRATIC CONVERGENCE

The distance  $|v - v_{[0]}| = |v[n+1] - v[n]|$  is shown to depend on the input by initially noticing that (21) can be used to define a recursion of the type  $f_{u[n+1]}(v) = f_{u[n]}(v) + \phi[n+1]$ , with

$$\begin{aligned} \phi[n+1] &= \rho C F_S (v[n-1] - v[n]) \\ &+ \frac{\rho}{R} (u[n] - u[n+1]). \end{aligned} \quad (44)$$

By the Lagrange theorem, if  $\xi$  is such that  $v[n+1] \leq \xi \leq v[n]$ :

$$\begin{aligned} f_{u[n+1]}(v[n+1]) &= f_{u[n+1]}(v[n]) + f'_{u[n+1]}(\xi)(v[n+1] - v[n]) \\ &= f_{u[n]}(v[n]) + \phi[n+1] + f'_{u[n+1]}(\xi)(v[n+1] - v[n]), \end{aligned}$$

and, since  $f_{u[n+1]}(v[n+1]) = f_{u[n]}(v[n]) = 0$ ,

$$|v[n+1] - v[n]| = \left| \frac{\phi[n+1]}{f'_{u[n+1]}(\xi)} \right|. \quad (45)$$

In the diode clipper inequality  $f'_u(v) > 1$  holds for each  $v$ , hence substituting (44) in (45):

$$\begin{aligned} |v[n+1] - v[n]| &\leq \rho C F_S |v[n] - v[n-1]| \\ &+ \frac{\rho}{R} |u[n+1] - u[n]|. \end{aligned} \quad (46)$$

Now, unfolding this inequality along  $n$  steps,

$$\begin{aligned} |v[n+1] - v[n]| &\leq (\rho C F_S)^n |v[1] - v[0]| \\ &+ \frac{\rho}{R} \sum_{k=1}^n (\rho C F_S)^{k-1} |u[k+1] - u[k]|, \end{aligned} \quad (47)$$

and by choosing  $k$  where  $|u[k+1] - u[k]|$  is maximum,

$$\begin{aligned} |v[n+1] - v[n]| &\leq (\rho C F_S)^n |v[1] - v[0]| \\ &+ \frac{\rho}{R} \max_{1 < k < n} |u[k+1] - u[k]| \sum_{k=1}^n (\rho C F_S)^{k-1}. \end{aligned} \quad (48)$$

It is straightforward to check that

$$\frac{\rho}{R} \sum_{k=1}^n (\rho C F_S)^{k-1} = 1 - (\rho C F_S)^n, \quad (49)$$

implying that the previous inequality can be expressed as

$$\begin{aligned} |v[n+1] - v[n]| &\leq (\rho C F_S)^n |v[1] - v[0]| \\ &+ \{1 - (\rho C F_S)^n\} \max_{1 < k < n} |u[k+1] - u[k]|. \end{aligned} \quad (50)$$



Finally, since the term  $(\rho C F_S)^n$  goes to zero for increasing  $n$ , it is convenient to rewrite the inequality as in (24).

## REFERENCES

- [1] J. Szczupak and S. K. Mitra, "Detection, location, and removal of delay-free loops in digital filter configurations," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. ASSP-23, no. 6, pp. 558–562, 1975.
- [2] A. Härmä, "Implementation of recursive filters having delay free loops," in *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP '98 (Cat. No.98CH36181)*, vol. 3, May 1998, pp. 1261–1264.
- [3] J. O. Smith, *Introduction to Digital Filters with Audio Applications*. <http://ccrma.stanford.edu/~jos/filters/>, Sep. 2007, online book.
- [4] S. Bilbao, *Wave and Scattering Methods for the Numerical Integration of Partial Differential Equations*. New York: John Wiley & Sons., 2004.
- [5] T. Kailath, *Linear Systems*. Englewood Cliffs: Prentice-Hall, 1980.
- [6] J. O. Smith *et al.*, *Spectral audio signal processing*. W3K, 2011, vol. 1334027739.
- [7] L. Trautmann and R. Rabenstein, *Digital sound synthesis by physical modeling using the functional transformation method*. Berlin: Springer, 2012.
- [8] V. Zavalishin, "The art of VA filter design," Native Instruments, Tech. Rep., 2012.
- [9] F. Avanzini and F. Fontana, "Exact discrete-time realization of a Dolby B encoding/decoding architecture," in *Proc. Conf. on Digital Audio Effects (DAFX-06)*, Montreal, Quebec, Canada, Sept. 18–20, 2006, pp. 297–302.
- [10] F. Fontana and M. Civolani, "Modeling of the EMS VCS3 voltage-controlled filter as a nonlinear filter network," *IEEE Trans. on Audio, Speech and Language Processing*, vol. 18, no. 4, pp. 760–772, 2010, special Issue on Virtual Analog Audio Effects and Musical Instruments.
- [11] F. Fontana and E. Bozzo, "Explicit fixed-point computation of nonlinear delay-free loop filter networks," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 10, pp. 1884–1896, Oct. 2018.
- [12] G. Borin, G. De Poli, and D. Rocchesso, "Elimination of delay-free loops in discrete-time models of nonlinear acoustic systems," *IEEE Trans. on Speech and Audio Processing*, vol. 8, no. 5, pp. 597–605, 2000.
- [13] D. T. Yeh, "Automated physical modeling of nonlinear audio circuits for real-time audio effects – Part II: Bjt and vacuum tube examples," *IEEE Trans. on Audio, Speech and Language Processing*, vol. 20, no. 4, pp. 1207–1216, May 2012.
- [14] M. Ducceschi, "A numerical scheme for various nonlinear forces, including collisions, which does not require an iterative root finder," in *Proc. Conf. on Digital Audio Effects (DAFX-17)*, Edinburgh, UK, Sep. 5-9 2017, pp. 80–86.
- [15] S. D'Angelo and V. Välimäki, "Generalized Moog ladder filter: Part I – Linear analysis and parameterization," *IEEE/ACM Trans. on Audio, Speech and Language Processing*, vol. 22, no. 12, pp. 1825–1832, Dec. 2014.
- [16] A. Huovilainen, "Nonlinear digital implementation of the Moog ladder filter," in *Proc. Conf. on Digital Audio Effects (DAFX-04)*, Naples, Italy, Oct. 2004, pp. 61–64.
- [17] K. Atkinson, *An Introduction to Numerical Analysis*. Wiley, 1989.
- [18] G. Evangelista, "Physical model of the string-fret interaction," in *Proc. Conf. on Digital Audio Effects (DAFX-11)*, Paris, Sep. 2011, pp. 345–351.
- [19] V. Chatziioannou, S. Schmutzhard, and S. Bilbao, "On iterative solutions for numerical collision models," in *Proc. Conf. on Digital Audio Effects (DAFX-17)*, Edinburgh, UK, Sep. 5-9 2017, pp. 72–79.

- [20] S. Papetti, F. Avanzini, and D. Rocchesso, "Numerical methods for a nonlinear impact model: a comparative study with closed-form corrections," *IEEE Trans. on Audio, Speech and Language Processing*, vol. 19, no. 7, pp. 2146–2158, 2011.
- [21] F. Avanzini, S. Serafin, and D. Rocchesso, "Interactive simulation of rigid body interaction with friction-induced sound generation," *IEEE Trans. on Speech and Audio Processing*, vol. 13, no. 5.2, pp. 1073–1081, 2005.
- [22] D. T. Yeh, J. S. Abel, A. Vladimirescu, and J. O. Smith, "Numerical methods for simulation of guitar distortion circuits," *Computer Music Journal*, vol. 32, no. 2, pp. 23–42, Summer 2008.
- [23] M. Karjalainen and J. Pakarinen, "Wave digital simulation of a vacuum-tube amplifier," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP2006)*. Toulouse, France: IEEE, May 15-19 2006, pp. 153–156.
- [24] I. Cohen and T. Hélie, "Simulation of a guitar amplifier stage for several triode models: examination of some relevant phenomena and choice of adapted numerical schemes," in *127th Convention of Audio Engineering Society*, New York, NY, Oct. 2009. [Online]. Available: <https://hal.archives-ouvertes.fr/hal-00631757>
- [25] J. Macak and J. Schimmel, "Real-time guitar preamp simulation using modified blockwise method and approximations," *EURASIP Journal on Advances in Signal Processing*, vol. 2011, no. 1, p. 629309, Feb 2011. [Online]. Available: <https://doi.org/10.1155/2011/629309>
- [26] M. Holters and U. Zölzer, "Physical modelling of a wahwah effect pedal as a case study for application of the nodal DK method to circuits with variable parts," in *Proc. Digital Audio Effects (DAFx-11)*, Paris, France, 2011, pp. 31–35.
- [27] F. Eichas, M. Fink, M. Holters, and U. Zölzer, "Physical modeling of the MXR Phase 90 Guitar Effect Pedal," in *Proc. Digital Audio Effects (DAFx-14)*, Erlangen-Nürnberg, Germany, 2014, pp. 153–158.
- [28] J. Chabassier, A. Chaigne, and P. Joly, "Modeling and simulation of a grand piano," *J. of the Acoustical Society of America*, vol. 134, no. 1, pp. 648–665, 2013.
- [29] A. Torin and M. Newton, "Collisions in drum membranes: a preliminary study on a simplified system," in *Proc. of the International Symposium on Musical Acoustics (ISMA 2014)*, Le Mans, France, Jul. 7 2014, pp. 401–406.
- [30] S. Zambon and F. Fontana, "Efficient polynomial implementation of the EMS VCS3 filter model," in *Proc. Conf. on Digital Audio Effects (DAFX-11)*, Paris, France, Sep. 2011, pp. 287–290.
- [31] J. Zhang and J. O. Smith III, "Real-time wave digital simulation of cascaded vacuum tube amplifiers using modified blockwise method," in *Proc. Conf. on Digital Audio Effects (DAFX-18)*, Aveiro, Portugal, Sep. 4-8 2018, pp. 141–148.
- [32] R. Muller and T. Hélie, "Power-balanced modelling of circuits as skew gradient systems," in *Proc. Conf. on Digital Audio Effects (DAFX-18)*, Aveiro, Portugal, Sep. 4-8 2018, pp. 264–271.
- [33] M. Holters and U. Zölzer, "Automatic decomposition of non-linear equation systems in audio effect circuit simulation," in *Proc. Conf. on Digital Audio Effects (DAFX-17)*, Edinburgh, UK, Sep. 5-9 2017, pp. 138–144.
- [34] J. Bridges and M. Van Walstijn, "Modal based tanpura simulation: Combining tension modulation and distributed bridge interaction," in *Proc. Conf. on Digital Audio Effects (DAFX-17)*, Edinburgh, UK, Sep. 5-9 2017, pp. 299–306.
- [35] T. Schwerdtfeger and A. Kummert, "A multidimensional approach to wave digital filters with multiple nonlinearities," in *2014 22nd European Signal Processing Conference (EUSIPCO)*, Sep. 2014, pp. 2405–2409.
- [36] K. J. Werner, V. Nangia, J. O. S. III, and J. S. Abel, "Resolving wave digital filters with multiple/multiport nonlinearities," in *Proc. Conf. on Digital Audio Effects (DAFX-15)*, Trondheim, Norway, Nov. 2015, pp. 387–394.
- [37] T. Schwerdtfeger and A. Kummert, "Newton's method for modularity-preserving multidimensional wave digital filters," in *2015 IEEE 9th International Workshop on Multidimensional (nD) Systems (nDS)*, Sep. 2015, pp. 1–6.
- [38] A. Bernardini, K. J. Werner, P. Maffezzoni, and A. Sarti, "Wave digital modeling of the diode-based ring modulator," in *Audio Engineering Society Convention 144*. Milan, Italy: Audio Engineering Society, May 2018, conv. paper #10015. [Online]. Available: <http://www.aes.org/e-lib/browse.cfm?elib=19411>

- [39] K. J. Werner, V. Nangia, A. Bernardini, J. O. Smith III, and A. Sarti, "An improved and generalized diode clipper model for wave digital filters," in *Audio Engineering Society Convention 139*. Audio Engineering Society, Oct. 2015. [Online]. Available: <http://www.aes.org/e-lib/browse.cfm?elib=17918>
- [40] R. Hoffmann-Burchardi, "Digital simulation of the diode ring modulator for musical applications," in *Proc. Conf. on Digital Audio Effects (DAFX-08)*, Espoo, Finland, Sep. 2008, pp. 165–168.
- [41] M. J. Olsen, K. J. Werner, and J. O. Smith, "Resolving grouped nonlinearities in wave digital filters using iterative techniques," in *Proc. Conf. on Digital Audio Effects (DAFX-16)*, Brno, Czech Republic, Sep. 2016, pp. 279–286.
- [42] D. Schleicher, "On the number of iterations of newton's method for complex polynomials," *Ergodic Theory and Dynamical Systems*, vol. 22, no. 3, p. 935945, 2002.
- [43] J. Hubbard, D. Schleicher, and S. Sutherland, "How to find all roots of complex polynomials by newton's method," *Inventiones Mathematicae*, vol. 146, no. 1, pp. 1–33, Oct. 2001.
- [44] G. Strang, *Introduction to Linear Algebra*. Wellesley, MA: Wellesley-Cambridge Press, 2016.
- [45] R. A. Brualdi and D. M. Cvetković, *A Combinatorial Approach to Matrix Theory and Its Applications*. Boca Raton, FL, USA: CRC Press, 2009.
- [46] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed. New York, NY, USA: Cambridge University Press, 2012.
- [47] F. Fontana, "Computation of linear filter networks containing delay-free loops, with an application to the waveguide mesh," *IEEE Trans. on Speech and Audio Processing*, vol. 11, no. 6, pp. 774–782, Nov. 2003.
- [48] P. Leni, Y. D. Fougere, and F. Truchetet, "Kolmogorov superposition theorem and its application to multivariate function decompositions and image representation," in *2008 IEEE International Conference on Signal Image Technology and Internet Based Systems*, Nov. 2008, pp. 344–351.

PLACE  
PHOTO  
HERE

**Federico Fontana** (SM11) received the Laurea degree in electronic engineering from the University of Padova, Italy, in 1996 and the Ph.D. degree in computer science from the University of Verona, Italy, in 2003. During the Ph.D. degree studies, he was a Research Consultant in the design and realization of real-time audio DSP systems. He is currently an Associate Professor in the Department of Mathematics, Computer Science and Physics, University of Udine, Italy, teaching Auditory & tactile interaction and Computer architectures. In 2001, he was Visiting Scholar at the Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology, Espoo, Finland. His current interests are in interactive sound processing methods and in the design and evaluation of musical interfaces. Professor Fontana coordinated the EU project 222107 NIW under the FP7 ICT-2007.8.0 FET-Open call from 2008 to 2011. Since 2017, he is Associate Editor of the IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING.

PLACE  
PHOTO  
HERE

**Enrico Bozzo** received the Laurea degree in computer science from the University of Udine, Italy, in 1990 and the Ph.D. degree in computer science from the University of Pisa, Italy, in 1994. He is currently an Assistant Professor in the Department of Mathematics, Computer Science and Physics, University of Udine, teaching numerical analysis. He was a team member in several national research projects. His current interests are in numerical linear algebra, in particular matrix theory and its applications.