STABLE STRUCTURES FOR NONLINEAR Biquad FILTERS

Jatin Chowdhury
Center for Computer Research in Music and Acoustics
Stanford University
Palo Alto, CA
jatin@ccrma.stanford.edu

ABSTRACT

Biquad filters are a common tool for filter design. In this writing, we develop two structures for creating biquad filters with nonlinear elements. We provide conditions for the guaranteed stability of the nonlinear filters, and derive expressions for instantaneous pole analysis. Finally, we examine example filters built with these nonlinear structures, and show how the first nonlinear structure can be used in the context of analog modelling.

1. INTRODUCTION

A “biquad” filter refers to a general 2nd order IIR filter. In digital signal processing, biquad filters are often useful since any higher-order filter can be implemented using a cascade of biquad filters. While digital biquad filters are typically implemented as linear processors, for audio applications it can be useful to implement nonlinear filters. For example, previous works have developed nonlinear simulations of analog audio circuits including Sallen-Key filters [1, 2], the Moog ladder filter [3, 4, 5], and more [6]. More relevant to our current topic is [7], in which the author suggests a method for altering a general digital feedback filter by saturating the feedback path, with the goal of achieving a more analog-like response. Finally, in [8], the author inserts dynamic range limiters into the biquad filter, with the intention of creating a hybrid limiter/filter effect. In this writing, we strive to develop more general nonlinear filter structures. While these structures may be used for analog modelling, they do not necessarily depend on analog modelling principles to be understood and implemented.

2. STRUCTURAL ELEMENTS

2.1. Linear Filter

We begin with the equation for a biquad filter

\[ y[n] = b_0 u[n] + b_1 u[n-1] + b_2 u[n-2] - a_1 y[n-1] - a_2 y[n-2] \] (1)

where \( y \) is the output signal, \( u \) is the input signal, and \( a_n \) and \( b_n \) are the feed-back and feed-forward filter coefficients, respectively. There are several convenient “direct forms” for implementing biquad filters. In this writing we will focus on the “Transposed Direct Form II” (TDF-II), which is popular for its favorable numerical properties [9]. Note that the poles of the filter can be described using the quadratic equation.

\[ p = \frac{-a_1 \pm \sqrt{a_1^2 - 4a_2}}{2} \] (2)

Specifically, the pole magnitude is described by (ignoring the trivial case where the poles are strictly real)

\[ |p|^2 = a_2 \] (3)

and the angular frequencies of the poles are equal to.

\[ \angle p = \arctan \left( \pm \frac{\sqrt{a_2 - a_1^2}}{a_1} \right) \] (4)

It is well known that a digital filter will be stable provided that the magnitudes of the poles are strictly less than 1 [9].

2.1.1. State-Space Formulation

Another reason TDF-II is useful for implementing biquad filters is that its behavior can easily be analyzed in state-space form. The state-space formalism is commonly used for constructing nonlinear virtual analog systems; brief introductions can be found in [10, 11]. To write the TDF-II biquad filter in state-space form, two state variables are defined at the locations of the delay elements.

\[ \begin{align*}
  x_1[n] &= b_1 u[n] - a_1 y[n] + x_2[n-1] \\
  x_2[n] &= b_2 u[n] - a_2 y[n]
\end{align*} \] (5)

Then the output of the filter can be written in terms of the states as,

\[ y[n] = b_0 u[n] + x_1[n-1] \] (6)
finally, the filter equation is written in a state-space form.

\[
\begin{bmatrix}
 x_1[n+1] \\
 x_2[n+1] \\
 y[n+1]
\end{bmatrix} =
\begin{bmatrix}
 0 & 1 & -a_1 \\
 0 & 0 & -a_2 \\
 1 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
 x_1[n] \\
 x_2[n] \\
 y[n]
\end{bmatrix} +
\begin{bmatrix}
 b_1 \\
 b_2 \\
 b_0
\end{bmatrix} u[n]
\]

(7)

2.2. Nonlinear Elements

We now propose adding nonlinear elements to the above filter structure. We will refer to these nonlinear elements as “base nonlinearities”. To keep the discussion as broad as possible, we consider any one-to-one nonlinear function \( f_{NL}(x) \).

In analog modelling literature, it is typical to analyze a nonlinear system by “linearizing” the system about a certain operating point. This process is typically done by constructing a Thévenin or Norton equivalent circuit that represents the nonlinear function at that operating point, where the resistance of the equivalent circuit is determined by the slope of the nonlinearity at the operating point, and the source of the equivalent circuit is determined by the DC offset of the linearized system at the operating point [12, 13].

In our purely digital formulation, we can linearize a nonlinear function as a gain element plus a constant source (see fig. 2). For example, the operating point \( x_0 \)

\[
f_{NL}(x) = f'_{NL}(x_0) x + c(x_0)
\]

(8)

where the offset \( c(x_0) \) is described by

\[
c(x_0) = f_{NL}(x_0) - f'_{NL}(x_0) x_0
\]

(9)

In fig. 3, we show an example of linearizing the nonlinear function \( f_{NL}(x) = \tanh(x) \) at \( x_0 = 1 \).

2.2.1. Stability Constraints

In order to guarantee that the filter structures described in the next section will be stable, we propose the following constraint on the base nonlinearities used to construct nonlinear filters:

\[
|f'_{NL}(x)| \leq 1
\]

(10)

In other words, the nonlinearities must never have a slope greater than 1. Many typical musical nonlinearities satisfy this constraint, including many saturating, dropout, rectifying, and warping nonlinearities. Note that this property is not satisfied by nonlinear functions that have discontinuous derivatives, such as \( f_{NL}(x) = |x| \). For functions of this type, we recommend using a smoothing scheme, such as BLAMP [14], to achieve a continuous first derivative.

Of particular interest to us will be saturating nonlinearities, including hard-clippers, soft-clippers, and sigmoid-like functions (see fig. 4). Saturating nonlinearities satisfy the property that

\[
|x| \rightarrow \infty, \quad f'_{sat}(x) \rightarrow 0
\]

(11)

![Comparing Saturating Nonlinearities](image)

2.3. Lyapunov Stability

As mentioned earlier, we can easily tell if a linear system is stable by analyzing the pole locations. For nonlinear systems, we need a more robust tool for analyzing stability; in this writing, we use Lyapunov stability [15], as has previously been applied to nonlinear digital waveguide networks [16], simulations of the Moog ladder filter [17], and direct form filters subject to fixed-point quantization [18]. To demonstrate that a system is Lyapunov stable, we must form the discrete-time state-space equation of the system

\[
x[n+1] = f(x[n])
\]

(12)

If every element of the Jacobian matrix of \( f \) is less than 1, at some operating point of the system, then the system is considered Lyapunov stable about that point. As discussed in [17], note that Lyapunov stability is a sufficient but not strictly necessary condition for the more general bounded-input, bounded-output (BIBO) stability.
3. NONLINEAR FILTER STRUCTURE 1: NONLINEAR BIQUAD

![Diagram of Nonlinear Transposed Direct Form II]

Figure 5: Nonlinear Transposed Direct Form II. The “NL” blocks refer to a generalized nonlinear element.

We now propose adding nonlinear elements to the TDF-II structure in the following fashion (see fig. 5). We will refer to this structure as the “Nonlinear Biquad”. The equation for the nonlinear biquad filter then becomes

\[
y[n] = b_0 u[n] + f_{NL}(b_1 u[n] - a_1 y[n-1] + f_{NL}(b_2 u[n-2] - a_2 y[n-2]))
\]

(13)

Here it can be useful to define the inputs to the nonlinearities.

\[
\begin{align*}
\chi_1 &= f_{NL}(\chi_2) + b_1 u[n-1] - a_1 y[n-1] \\
\chi_2 &= b_2 u[n-2] - a_2 y[n-2]
\end{align*}
\]

(14)

Note that for saturating base nonlinearities, as the input \( u \) grows large, the other terms will become negligible.

Now we can replace the nonlinear elements with their linearized models, using the state variables to define the operating points. To make our notation more concise, we will denote the output of the nonlinear functions as follows

\[
f_{NL}(\chi_k) = g_k \chi_k + \gamma_k
\]

(15)

Then eq. (13) can be re-written

\[
y[n] = b_0 u[n] + g_1 (b_1 u[n-1] - a_1 y[n-1]) + g_2 (b_2 u[n-2] - a_2 y[n-2]) + \gamma_1
\]

(16)

Finally, we can re-write the filter coefficients as variables dependent on the state variables.

\[
\begin{align*}
b_0' &= b_0 \\
b_1' &= g_1 b_1 \\
b_2' &= g_1 g_2 b_2 \\
a_1' &= g_1 a_1 \\
a_2' &= g_1 g_2 a_2
\end{align*}
\]

(17)

\[
y'[n] = b'_0 u[n] + b'_1 u[n-1] - a'_1 y'[n-1] + b'_2 u[n-2] - a'_2 y'[n-2] + g_1 \gamma_2 + \gamma_1
\]

(18)

Note that the two \( \gamma \) terms in eq. (18) are simple offsets as defined by our linearized model, and as such will not affect the pole locations, nor the filter stability.

3.1. Stability

Recall that the linear biquad filter equation can be written in state-space form as in eq. (7). By writing the nonlinear biquad eq. (13), in the state-space form defined by eq. (12), we find

\[
\begin{bmatrix}
x_1[n+1] \\
x_2[n+1] \\
y'[n+1]
\end{bmatrix} = h \begin{bmatrix} x_1[n] \\ x_2[n] \\ y'[n] \end{bmatrix} + b \begin{bmatrix} b_1 \\ b_2 \\ b_0 \end{bmatrix} u[n]
\]

(19)

where,

\[
\begin{align*}
h_1(x_1[n], x_2[n], y[n]) &= f_{NL}(x_2[n]) - a_1 y[n] \\
h_2(x_1[n], x_2[n], y[n]) &= -a_2 y[n] \\
h_3(x_1[n], x_2[n], y[n]) &= f_{NL}(x_1[n])
\end{align*}
\]

(20)

then the Jacobian matrix of \( h \) can be written as follows.

\[
J = \begin{bmatrix}
0 & f'_{NL}(x_2[n]) & -a_1 \\
0 & 0 & -a_2 \\
0 & 0 & 0
\end{bmatrix}
\]

(21)

From this analysis, we see that the nonlinear biquad filter will be stable, provided that the constraint from eq. (10) is satisfied, and all the \( a \) coefficients are less than 1. However, note that the constraint on the \( a \) coefficients is required anyway for the corresponding linear filter to be stable, so the only “new” constraint that arises from adding the nonlinear elements is that of eq. (10).

3.2. Pole Analysis

Since the coefficients of the biquad filter will be dependent on the state of the filter, the instantaneous poles of the filter will be dependent as well. In order to calculate the instantaneous poles of the nonlinear biquad structure, we can adjust the formula from eq. (2).

\[
p' = -g_1 a_1 \pm \sqrt{g_1^2 a_1^2 - 4 g_1 g_2 a_2} \over 2
\]

(22)

For saturating base nonlinearities, we can see from eq. (11) that as the state variables grow large, the poles will go to zero.

The pole magnitude and angle with move as follows.

\[
|p|^2 = g_1 g_2 a_2
\]

(23)

\[
\angle p = \arctan \left( \pm \sqrt{\frac{g_2 a_2 - a_1^2}{a_1}} \right)
\]

(24)

Note that while the two gain elements \((g_1, g_2)\) are approximately equal, the nonlinear pole will have the same angle as the corresponding linear pole. An example of this pole movement can be seen in fig. 6: note that as the state variables grow large and the \( g \) factors become smaller, the poles and zeros both approach zero.
4. NONLINEAR FILTER STRUCTURE 2: NONLINEAR FEEDBACK FILTER

We now propose a different structure for adding elements to a TDF-II Biquad filter, this time adding nonlinear elements to the feedback paths (see Fig. 7). Note that although the two structures are developed separately here, they could certainly be combined into a third structure, which will also be stable under the same conditions as the two original structures. The equation for the filter can now be written.

\[
y[n] = b_0 u[n] + b_1 u[n-1] + b_2 u[n-2] - a_1 f_{NL}(y[n-1]) - a_2 f_{NL}(y[n-2])
\]

(25)

Again, we can replace the nonlinear elements with their linearized models, this time using the \(y[n-1]\) and \(y[n-2]\) terms to define our operating points.

\[
f_{NL,k}(x) = g_k x + \gamma_k
\]

(26)

And again, the filter equation can be re-written as,

\[
y[n] = \hat{b}_0 u[n] + \hat{b}_1 u[n-1] + \hat{b}_2 u[n-2] - a_1 g_1 y[n-1] + \gamma_1 - a_2 (g_2 y[n-2] + \gamma_2)
\]

(27)

or by re-writing the filter coefficients, we see.

\[
\begin{align*}
\hat{b}_0 &= b_0 \\
\hat{b}_1 &= b_1 \\
\hat{b}_2 &= b_2 \\
a_1' &= g_1 a_1 \\
a_2' &= g_2 a_2
\end{align*}
\]

(28)

\[
y'[n] = b_0' u[n] + b_1' u[n-1] - a_1' y[n-1] + b_2' u[n-2] - a_2' y[n-2] - a_1 \gamma_1 - a_2 \gamma_2
\]

(29)

Again, the \(\gamma\) offset terms will not affect the filter stability.

4.1. Stability

We can now update eq. (7) for the nonlinear feedback filter described by eq. (25), and by writing it in the form of eq. (12), we see

\[
\begin{bmatrix}
x_1[n+1] \\
x_2[n+1] \\
y'[n+1]
\end{bmatrix} = h \begin{bmatrix} x_1[n] \\
x_2[n] \\
y[n]
\end{bmatrix} + \begin{bmatrix} b_1 \\
b_2 \\
b_0
\end{bmatrix} u[n]
\]

(30)

where,

\[
\begin{align*}
h_1(x_1[n], x_2[n], y[n]) &= x_2[n] - a_1 f_{NL}(y[n]) \\
h_2(x_1[n], x_2[n], y[n]) &= -a_2 f_{NL}(y[n]) \\
h_3(x_1[n], x_2[n], y[n]) &= x_1[n]
\end{align*}
\]

(31)

then the Jacobian matrix can be written as follows.

\[
J = \begin{bmatrix}
0 & 1 & -a_1 f_{NL}'(y[n]) \\
0 & 0 & -a_2 f_{NL}'(y[n]) \\
1 & 0 & 0
\end{bmatrix}
\]

(32)

Again, assuming that the corresponding linear filter is stable, the nonlinear feedback filter will be stable provided the constraint from eq. (10) is satisfied, that the absolute value of \(f_{NL}'(x)\) is always less than or equal to 1.

4.2. Pole Analysis

We can now calculate the locations of the instantaneous poles for the nonlinear feedback filter, by adjusting the formula from eq. (2).

\[
p' = -g_1 a_1 \pm \sqrt{a_1^2 + 4a_2 g_2}
\]

(33)
Figure 8: Instantaneous poles for a resonant lowpass filter with nonlinear feedback, calculated from eq. (33), with $g_1 = g_2 = g$.

In this case the pole magnitude and angle will move as follows.

$$|p|^2 = g_2 a^2$$  \hfill (34)

$$\angle p = \arctan \left( \pm \frac{\sqrt{g_2^2 a^2 - a_1^2}}{a_1} \right)$$  \hfill (35)

Note that for saturating nonlinearities, the pole magnitude decays to zero more slowly than for the nonlinear biquad. More importantly, as the input gain increases, the pole angle increases as well, creating a sonically interesting “sweeping” sound. [5] describes this sort of pole movement as “audio-rate modulation of the cutoff” for the filter, which can be a useful way of thinking about this phenomenon. This pole movement is particularly noticeable in the biquad structure created in [8], which can be seen as a special case of the nonlinear feedback filter described here.

Finally, note that unlike the nonlinear biquad, the zeros of the filter are not affected by the nonlinear elements. While adding nonlinear elements to the feedforward path can introduce a similar effect for the zeros, this would be functionally equivalent to processing the signal through a nonlinearity before passing it into the filter. Again, an example of this pole movement can be seen in fig. 8: note that as the state variables grow large and the $g$ factors become smaller, the poles angles increase and the pole magnitude shrinks to zero, while the zeros remain unchanged.

5. EXAMPLE: RESONANT LOWPASS FILTER

As an example of the nonlinear structures developed above, we will now examine a resonant lowpass filter designed with both nonlinear structures. We will then show how the nonlinear biquad structure can be useful for analog modelling, and compare to an analog filter made with the same specifications.

Our example filter will be a lowpass filter with a cutoff frequency at $f_c = 1$ kHz, and $Q = 10$. For our nonlinear elements, we will use a hyperbolic tangent function $f_{NL}(x) = \tanh(x)$. Note that this nonlinear function belongs to the class of saturating nonlinearities described by eq. (11).

5.1. Digital Nonlinear Biquad

We first construct this filter using the nonlinear biquad structure. In fig. 9 we show the frequency-domain output of this filter for sine sweeps of various amplitudes, compared to the frequency response of the corresponding linear filter. Note that “frequency-response” is an ill-defined concept for nonlinear systems; as a result, these plots should be seen as a roughly approximating the frequency response at particular operating points, rather than true frequency response measurements. In fig. 6 we show the movement of the poles and zeros of the filter for varying steady state inputs. We calculate the instantaneous poles using eq. (22), using $g_1 = g_2 = g$, as described in each figure.

5.2. Digital Nonlinear Feedback Filter

Next, we construct the same resonant lowpass filter using the nonlinear feedback structure. In fig. 10, we show the sine-sweep response of the filter at various amplitudes. In fig. 8 we show the movement of the poles and zeros of the filter for varying steady state gains. The instantaneous poles are calculated using eq. (33), again using $g_1 = g_2 = g$.

5.3. Using the Nonlinear Biquad for Analog Modelling

To show how the nonlinear biquad filter structure can be useful for analog modelling purposes, first note that the input gain to the nonlinear biquad can be used as a tunable parameter (see fig. 11).
By tuning the input gain, we can attempt to match the response of an arbitrary analog filter, either by tuning the parameters by ear or using some form of numerical optimisation. Note that the choice of base nonlinearities used by the nonlinear biquad will also play a role in the accuracy of the model. For example, if the analog filter being modelled displays asymmetric nonlinear properties, then in order to accurately model that filter, the nonlinear biquad must be constructed using asymmetric base nonlinearities.

5.3.1. Comparison with Analog Filter

As an example, we can attempt to construct a naive model of a Sallen-Key lowpass filter, a commonly used analog filter structure, and compare our results to the desired analog response, similar to the comparison done in [1]. We describe this as a naive model because we do not make any attempt to understand the physical properties of the analog filter when constructing this model. We construct a nonlinear biquad filter using tanh base nonlinearities, and design a resonant lowpass filter with cutoff frequency $f_c = 1$ kHz, and $Q = 10$, as well as a simulation of the corresponding Sallen-Key filter using LTSpice. To accentuate the nonlinear behavior of the analog filter, we choose ±4 V as the source voltages for the analog filter circuit.

We then compare the outputs of the two filters for square waves at different frequencies, and use a simple staircase optimisation scheme to find the input gain for the nonlinear biquad that best matches the analog simulation. The results for the 250 Hz square wave can be seen in fig. 12. While the nonlinear biquad model is far from perfect, it does capture the damping effects of the analog filter much more accurately than the corresponding linear filter, and could be further improved with a more well-informed choice of base nonlinear functions, and a more sophisticated optimisation scheme.

6. CONCLUSION

In this paper, we have developed two structures for stable nonlinear biquad filters: the “nonlinear biquad filter” and “nonlinear feedback filter.” We have introduced the new architectures as a modification of the Transposed Direct Form II filter structure, and shown how the changed architectures affect the pole locations depending on the amplitude of the input signal. We have also derived constraints under which the structures are guaranteed stable.

As a case study, we have implemented a resonant lowpass filter using both nonlinear structures, and shown that the poles respond to the input as expected. We have also shown that the nonlinear biquad structure can be used to model analog filters, comparing with a Sallen-Key lowpass filter as an example. Note that while the nonlinear biquad structure can be used for analog modelling, both structures can also be used purely in the digital domain as a tool for constructing filters that sound more sonically interesting and harmonically rich.

To demonstrate this last point, we have also developed an open-source audio plugin (VST, AU) implementation of both the nonlinear biquad and nonlinear feedback filters, extending to several filter shapes, and several base nonlinearities. The source code for the
plugin implementation is available on GitHub,¹ and video demonstrations are available on YouTube.²,³

Future research on this topic will center around making better-informed choices of base nonlinearities, focusing on both the desired harmonic response of the filter, as well as physically meaningful base nonlinearities for use in analog modelling. Another possible area of inquiry involves using these filter structures to model the TR-808 bass drum circuit, which contains a resonant filter that exhibits level-dependent cutoff-frequency modulation similar to the nonlinear feedback filter [19]. Finally, the filter structures developed in this writing could be used to design time-varying amplitude-dependent filters, e.g. the time-varying modal filters described in [20].

7. ACKNOWLEDGMENTS

The author would like to thank Kurt Werner for providing sound advice and guidance on this topic, Viraga Perera and Dave Berners for informative discussions, as well as the GASP working group.

8. REFERENCES


¹https://github.com/jatinchowdhury18/ComplexNonlinearities
²https://youtu.be/BMzKda2tmoU
³https://youtu.be/T0ASIX5oL9A