Computational optimization of nonlinear zero-delay feedback by second-order piecewise approximation

Vadim Zavalishin*

November 7, 2008

Abstract

The article discusses piecewise approximation of nonlinear shaping curves by second-order segments. For such approximations, the nonlinear zero-delay feedback equation of a discrete-time LTI system can be solved analytically. Two application examples are given, using an approximation of a hyperbolic tangent-based nonlinearity in the ladder and state-variable filters respectively.

1 Introduction

The computation of nonlinearities in zero delay feedback systems is generally somewhat problematic [1]. A special case however occurs, if the feedback path contains a single nonlinear memoryless shaper $y = S(x)$ which can be implicitly represented as a second-order curve $S(x, y) = 0$:

$$ax^2 - 2bxy + cy^2 - 2px - 2qy + r = 0 \tag{1}$$

The above can be considered as a quadratic equation relative to y, having the form:

$$Ay^2 - 2By + C = 0 \tag{2}$$

where

$$\begin{aligned} A &= c \\ B &= bx + q \\ C &= ax^2 - 2px + r \end{aligned} \tag{3}$$

*Native Instruments GmbH

Solving (2), we obtain an explicit form for $y = S(x)$:

$$y = \frac{B + \sigma\sqrt{B^2 - AC}}{A} \tag{4a}$$

$$= \frac{C}{B - \sigma\sqrt{B^2 - AC}} \tag{4b}$$

where σ is taking *only one* of the values $\{-1, 1\}$. Apparently, the choice of σ selects one of the two branches of the second-oder curve defined by (1).

Notice that if $c = 0$, then (1) turns into a rational function

$$y = \frac{1}{2} \cdot \frac{ax^2 - 2px + r}{bx + q} \tag{5}$$

Now, a general zero-delay feedback equation for the case of a single contained nonlinearity $y = S(x)$ has the form:

$$\mu t + \nu = S(\alpha t + \beta) \tag{6}$$

where β and ν depend on the system state and system parameters, α and μ depend on the system parameters only [1], and t is an unknown signal at some point in the system structure. The same can obviously be written in the implicit form $S(x, y) = 0$ as:

$$S(\alpha t + \beta, \mu t + \nu) = 0 \tag{7}$$

Letting

$$\begin{aligned} x &= \alpha t + \beta \\ y &= \mu t + \nu \end{aligned} \tag{8}$$

and substituting (8) into (1) we obtain a quadratic equation against t, which can be solved analytically using formulas similar to (4).[1]

The important point of the above is, that in this case the nonlinear feedback equation can be solved analytically, without resorting to iterative algoritms.

2 Piecewise approximation

Considering the spectrum of curves defined by (1), we might notice that probably none of the typical nonlinearity shapes used in LTI systems can be reasonably approximated by (4). We can however hope to achieve a reasonable result if we don't try to approximate the entire nonlinearity $y = S(x)$ by (4), but rather split the x axis into consecutive segments $[x_i, x_{i+1}]$, each using its own approximation of the form (4). It's important to notice, that (as will become clear later in the text) there is no special benefit if the segments have equal lengths.

We will therefore formulate our problem as follows: for a given target function $y = f_T(x)$ and a given N, find a set of N segments $[x_i, x_{i+1}]$, covering the entire real axis[2], and a corresponding set of approximating functions $y = f_i(x)$, where each $f_i(x)$ has the form (4) and is defined on $[x_i, x_{i+1}]$. We want to find such set of $[x_i, x_{i+1}]$ and corresponding $f_i(x)$, that the resulting approximation $f(x)$:

$$f(x) = f_i(x) \qquad x \in [x_i, x_{i+1}]$$

is 'reasonably close' to $f_T(x)$.

As a measure of 'being close' we might chose the absolute error:

$$\varepsilon = \sup |f(x) - f_T(x)|$$

or the relative error:

$$\varepsilon = \sup \left| \frac{f(x)}{f_T(x)} - 1 \right|$$

In this article, we will use the relative error, to achieve a good approximation for low level signals, where the distortion caused by the nonlinearity should be generally low.

Ideally we would want to find a solution which minimizes ε, which is not an easy task. Fortunately, getting reasonably close to the optimal solution often should be satisfactory. In this article we will consider a number of approaches to find such reasonably close approximations.

Apparently, the algorithm for solving the equation (6), outlined earlier in the article, cannot be immediately applied to a piecewise function. We can however notice that the equation (6) can be graphically interpreted as an intersection of a straight line (8) with the curve $y = S(x)$.

Expressing (8) in the implicit form we get:

$$(x - \beta)\mu = (y - \nu)\alpha \tag{9}$$

from where the (non-normalized) signed distance from a given point (x, y) to the line can be computed as

$$d(x, y) = (x - \beta)\mu - (y - \nu)\alpha$$

Now, assume the coefficients α, β, μ, ν are such that there is only one intersection of (9) with the approximating curve $y = f(x)$. Also, let $y_i = f(x_i)$. Then we can find out, which of the segments $f_i(x)$ is intersecting with (9), by checking the signed distances $d_i = d(x_i, y_i)$. Obviously, under given conditions, there will be only one i such that the points (x_i, y_i) and (x_{i+1}, y_{i+1}) lie on the different sides of (9).

Furthermore, often the sequence of signed distances d_i will be monotonic, in which case the intersecting segment can be found by a simple binary search.

3 Auxiliary devices

3.1 Odd continuation

Often the nonlinearity $S(x)$ will be an odd function. In that case we don't have to approximate $S(x)$ on the whole real axis, rather we can build the approximation on $[0, +\infty]$. Assuming $f_i(x)$ is defined on

[1]The solution forms (4a) and (4b) are theoretically equal, in practice however we should make a choice based on the sign of σB, if we are to avoid the unnecessary precision loss (or even a division by zero).

[2]Since $[x_i, x_{i+1}]$ cover the entire real axis, it follows that $x_0 = -\infty$, $x_N = +\infty$.

some $[x_i, x_{i+1}]$ where $0 \le x_i < x_{i+1} \le +\infty$ and has the form (1), we can easily find out it's odd counterpart $\bar{f}_i(x) = -f_i(-x)$ defined on $[-x_{i+1}, -x_i]$ by letting $\bar{a} = a$, $\bar{b} = b$, $\bar{c} = c$, $\bar{p} = -p$, $\bar{q} = -q$, $\bar{r} = r$, $\bar{\sigma} = -\sigma$.

Thus, in this case we can first solve the approximation problem for $x_0 = 0$, $x_N = +\infty$ and build the remaining segments as described above.

3.2 Warped coordinates

To be able to more easily handle the infinite segment $[0, +\infty]$ we might map it onto a finite segment. E.g. we could use the mapping

$$\tilde{x} = \frac{x}{x+1} \qquad x = \frac{\tilde{x}}{1-\tilde{x}} \tag{10}$$

which maps $[0, +\infty]$ onto $[0, 1]$.[3]

The warped coordinates allow us for example to compute the 'middle' of any segment $[x_i, x_{i+1}]$, or to compare the segment 'lengths'.

3.3 Interpolating segments

At this point we will define an algorithm which will serve as a basis for the approximation approaches described later. This algorithm is building an *interpolating segment* of the form (4) between two given points x_i and x_{i+1} of the curve $f_T(x)$.

Let

$$\begin{aligned} y_i &= f_T(x_i) \\ y_{i+1} &= f_T(x_{i+1}) \\ y'_i &= f'_T(x_i) \\ y'_{i+1} &= f'_T(x_{i+1}) \end{aligned} \tag{11}$$

We wish to build $f_i(x)$ of the form (4), which satisfies

$$\begin{aligned} f_i(x_i) &= y_i \\ f_i(x_{i+1}) &= y_{i+1} \\ f'_i(x_i) &= y'_i \\ f'_i(x_{i+1}) &= y'_{i+1} \end{aligned} \tag{12}$$

[3]Strictly speaking $\pm\infty$ maps to 1, but since the infinity should be anyway treated in a limiting sense, we can always ensure the positive sign thereof.

Apparently, (1) (or for that matter (4)) has 5 degrees of freedom, whereas (12) has only 4 equations. We will use the 5th degree of freedom to minimize the possible approximation error on this segment. While there is a number of different ways to define the 5th equation, we will use the following one:

$$f_i(x_{i+0.5}) = y_{i+0.5} \tag{13}$$

where $x_{i+0.5}$ is some point 'in the middle' of the segment and $y_{i+0.5}$ is the specified value of $f_i(x)$ at this point.[4] Thus $y_{i+0.5}$ is a free parameter and we can attempt to minimize the error by varying it. Usually there will be an error minimum somewhere around $y_{i+0.5} \approx f_T(x_{i+0.5})$, which can be found using e.g. the golden ratio search.

Substituting (12) and (13) into (1), and differentiating (1) against x where necessary, we obtain the following system:

$$\begin{aligned} & ax_i^2 - 2bx_iy_i + cy_i^2 - 2px_i - 2qy_i + r = 0 \\ & ax_{i+1}^2 - 2bx_{i+1}y_{i+1} + cy_{i+1}^2 \\ & \qquad - 2px_{i+1} - 2qy_{i+1} + r = 0 \\ & ax_i - b(y_i + x_iy'_i) + cy_iy'_i - p - qy'_i = 0 \\ & ax_{i+1} - b(y_{i+1} + x_{i+1}y'_{i+1}) + cy_{i+1}y'_{i+1} \\ & \qquad - p - qy'_{i+1} = 0 \\ & ax_{i+0.5}^2 - 2bx_{i+0.5}y_{i+0.5} + cy_{i+0.5}^2 \\ & \qquad - 2px_{i+0.5} - 2qy_{i+0.5} + r = 0 \end{aligned} \tag{14}$$

where the unknowns are a, b, c, p, q, r. Since the system is homogeneous, the solutions form a straight line going through the origin of a 6-dimensional space. Apparently, any single one of them (except the trivial one) will do.[5]

[4]The exact choice of the middle point $x_{i+0.5}$ is not too important, e.g. a simple way to choose it is to take the 'warped middle', that is the middle of the warped image of $[x_i, x_{i+1}]$.

[5]An easy way to solve (14) is obviously to let one of the unknowns equal to some predefined value, e.g. $c = 1$ or $q = 1$, which provides the 6th equation and makes the system inhomogeneous at the same time. Notice that this excludes the possibility of finding the solution line lying within the hyperplane $c = 0$ or $q = 0$ respectively (one could however expect to be able detect this case from having the determinant of the resulting 6-equation system equal to zero).

Having found the coefficients a, b, c, p, q, r, we still need to find the value of σ. To do that we can simply first try $\sigma = 1$ and $\sigma = -1$ in turn, while computing the values of $f_i(x)$ and $f'_i(x)$ at x_i and x_{i+1} and comparing those to the values specified by (12).[6]

Notice, that the just described algorithm is not limited to (11). In fact, we could use any other specification for the desired values of $f_i(x)$ and $f'_i(x)$ at x_i and x_{i+1}. One should however take care to specify only such points (x_i, y_i), (x_{i+1}, y_{i+1}) and the derivatives y'_i, y'_{i+1} (and $(x_{i+0.5}, y_{i+0.5})$ for that matter), that they can be connected by a curve without an inflection point. Otherwise, since the curve (1) cannot contain inflection points, we won't be able to choose σ which satisfies the conditions (12) at both ends of the segment simultaneously, or (4) will get discontinuous on the segment.

3.4 Interpolation at the infinity

Since the last segment of the piecewise approximation should extend to infinity, we have $x_N = +\infty$. Apparently, one cannot use the equation system (14) in that case, and we need to obtain another form for the conditions at x_N.

If the target curve $f_T(x)$ is having the 'saturation' behavior, then there is a horizontal asymptote at $x = +\infty$. So we let

$$\begin{aligned} y_{+\infty} &= f_T(+\infty) \\ y'_{+\infty} &= 0 \end{aligned}$$

which turns into the equations for the approximating function $f_i(x)$:

$$\begin{aligned} f_i(+\infty) &= y_{+\infty} \\ f'_i(+\infty) &= 0 \end{aligned} \tag{15}$$

Now apparently (4) can have a horizontal asymptote only if $a = 0$ and $b\sigma \leq 0$. Using (4b) we obtain

$$\lim_{x \to +\infty} f_i(x) = -p/b \qquad (b \neq 0)$$

therefore (15) can be expressed as

$$\begin{aligned} a &= 0 \\ by_{+\infty} + p &= 0 \end{aligned}$$

which should replace the respective equations in the system (14). One should also remember to check the condition $b\sigma < 0$ during the choice of σ.

4 Approximation approaches

The approaches we describe here will assume that $S(x)$ is odd and we are using the odd continuation. However, it should be easy to adapt them to the case of $x_0 = -\infty$, $x_N = +\infty$.

4.1 Regular interpolation

We can let

$$x_i = \begin{cases} i \cdot \Delta x & \text{for } i = 0, \ldots, N-1 \\ +\infty & \text{for } i = N \end{cases}$$

where Δx is the step. Given the above set of x_i, we can build the interpolating segments satisfying (11), as described earlier in the article.

Apparently, increasing Δx will generally increase the approximation error on $[x_0, x_{N-1}]$ and decrease it on $[x_{N-1}, x_N]$. Decreasing Δx will do the reverse. Therefore there generally should be a value of Δx where both errors are equal, and the error on the whole range $[0, +\infty]$ is having a local minimum. This value can be found using e.g. the golden ratio method.

If one wishes to have a more reliable method, they could bring (14) to the trapezoid form using the Gauss method, so that the last equation will contain only two unknowns, thus it should be easy to find a nontrivial solution.

Of course, the usage of more elaborate methods for solving homogeneous linear equation systems is not excluded.

[6] If $y = f(x)$ is having the form (4), then by differentiating (1) by x it is easy to obtain the following expression for $f'(x)$:

$$y' = -\frac{ax - by - p}{cy - bx + q}$$

Notice that the above expression implicitly contains σ by incorporating the value of y into the right side.

4.2 Warped-regular interpolation

In this method, we let

$$\tilde{x}_i = i/N$$

where $\tilde{x}$ is defined by (10). The segments are again built accordingly to (11).

4.3 Irregular interpolation

Assume we have a division of $[0, +\infty]$ defined by a given set of x_i. Let $f_i(x)$ be the approximating segments built accordingly to (11), and let ε_i be the approximation errors on those segments. Further, let ε be the average error on all segments (the exact formula for ε is not important, e.g. the arithmetic average of ε_i could do, provided all ε_i have a comparable order of magnitude), and let $\tilde{l}_i = \tilde{x}_{i+1} - \tilde{x}_i$ be the 'warped lengths' of the segments.

Then we can compute the new segment lengths according to

$$\tilde{l}'_i = \tilde{l}_i \cdot \left(\frac{\varepsilon}{\varepsilon_i}\right)^\alpha$$

where $\alpha \ll 1$ is a braking factor. Apparently the sum of the warped lengths should be equal to 1, so the final length replacement formula is

$$\tilde{l}_i \longleftarrow \tilde{l}'_i \Big/ \left(\sum \tilde{l}'_i\right)$$

The process should generally converge after a number of iterations, resulting in equal ε_i. The initial set of x_i can be e.g. built by using the warped-regular interpolation.

4.4 Further enhancements

As mentioned earlier, in building the interpolating segments we don't need to restrict ourselves to (11). It is still however probably a good idea, to keep the first derivative continuous at x_i.

Fig. 1 demonstrates the relative deviation for the case of two interpolating segments built according to (11). Alternatively, we could e.g. offset the end point specifications (11) as to get the picture in Fig. 2 (notice that the first derivative is still continuous in the latter version). This would generally reduce the approximation error, although not necessarily too significantly.

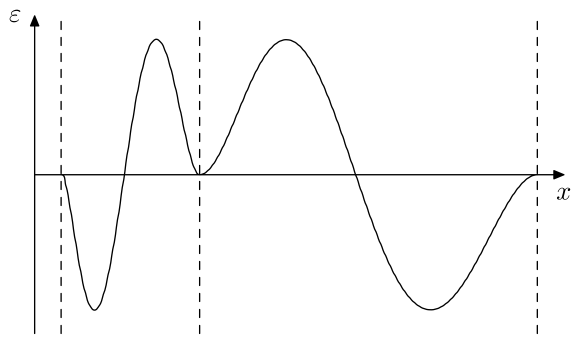

Figure 1: Connection of interpolating segments.

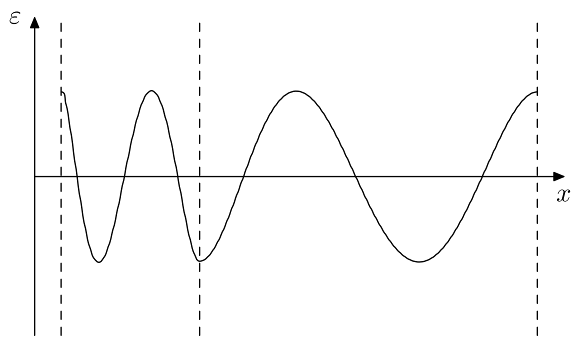

Figure 2: Alternative segment generation.

5 Approximating the tanh function

Since tanh is a commonly used saturation curve, we will build a piecewise approximation thereof, for the example purposes. Since the function is odd, we can use the odd continuation approach. Trying out a single segment on the positive x semiaxis, we have $N = 1$, $x_0 = 0$, $x_1 = +\infty$, $y_0 = 0$, $y'_0 = 1$, $y_1 = 1$, $y'_1 = 0$.

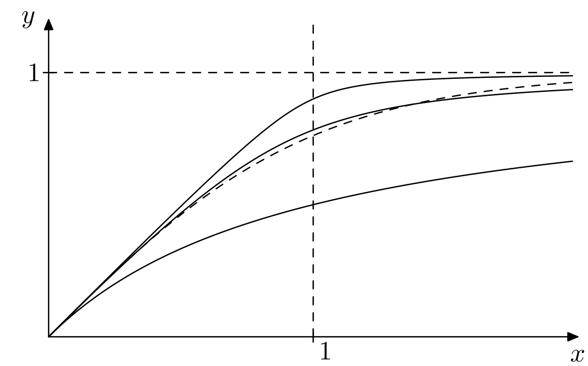

Figure 3: The family of curves defined by (16). The dashed curve is the tanh function.

Letting $q = 1$, $x_{0.5} = 1$, and solving (14) we obtain

$$\begin{aligned}&a = 0,\ b = 1,\ p = -1,\ q = 1,\ r = 0,\\ &\sigma = -1,\ c = \frac{2}{y_{0.5}} \cdot \left(2 - \frac{1}{y_{0.5}}\right)\end{aligned} \tag{16}$$

Optimizing the relative error against $y_{0.5}$ we obtain the minimum $\varepsilon \approx 0.033$ (which corresponds to approximately -30dB) at $y_{0.5} = 0.7829231$ and $c = 1.8462486$. Taking into account the fact that the obtained curve has a continuous 1st derivative, we might consider the result to be 'good enough' and not try larger values of N to improve the approximation precision.

Notice, that the curve shape of (16) practically offers itself to be morphed by varying $y_{0.5}$ within the interval $0 < y_{0.5} < 1$, which results in $-\infty < c < 2$ (Fig. 3). Particularly, at $y_{0.5} = 0.5$ we have $c = 0$ and the curve turns into a simple hyperbolic saturator $y = x/(1 + x)$.

Apparently, for $c < 2$ the curve (16) doesn't have discontinuities on $[0, +\infty)$. Therefore, since $f'(0) = 1$ and since there are no inflection points, $0 < f'(x) < 1$ for all $x > 0$.

6 Ladder filter

Now, for the example purposes, we are going to build a ladder filter model, containing the piecewise-approximated saturator at the feedback point. In

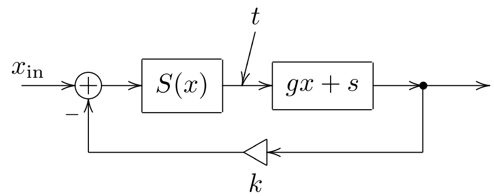

Figure 4: Ladder filter with a saturator.

Fig. 4, $S(x)$ is the saturator and $gx + s$ is the instant response [1] of the LTI subsystem denoted by the respective box, usually a serial chain of four 1-pole lowpass filters. Particularly, for a chain of bilinear-transformed 1-pole lowpass filters we have $0 < g < 1$, which is what we are going to assume further in this model. We are also going to assume $k > -1$ and $0 \le S'(x) \le 1$, so that there will be exactly one intersection between $S(x)$ and (8).

By choosing the signal at the output of the saturator as the unknown t, we obtain the equation (7) for this structure in the form:

$$S(x_{\text{in}} - k(gt + s), t) = 0 \tag{17}$$

that is

$$\alpha = -gk, \beta = x_{\text{in}} - ks, \mu = 1, \nu = 0$$

(notice that $\alpha < 1$).

The signed line distances are therefore

$$d_i = d(x_i, y_i) = x_i - \beta - \alpha y_i$$

Since $0 \le S'(x) \le 1$ and $\alpha < 1$, then (since $y_i = S(x_i)$) the sequence $\{d_i\}$ should be strictly ascending. Thus, the index i of the 'active segment' (that is the one participating in the intersection with the line) can be easily found by a binary search within $\{d_i\}$.

Substituting (17) into (1) and bringing that to the form (2), we obtain

$$A't^2 - 2B't + C' = 0 \tag{18}$$

where

$$\begin{aligned}A' &= a\alpha^2 - 2b\alpha + c\\ B' &= -a\alpha\beta + b\beta + p\alpha + q\\ C' &= a\beta^2 - 2p\beta + r\end{aligned} \tag{19}$$

and therefore

$$t = \frac{B' + \sigma'\sqrt{B'^2 - A'C'}}{A'} \tag{20a}$$

$$= \frac{C'}{B' - \sigma'\sqrt{B'^2 - A'C'}} \tag{20b}$$

To find out the value of σ' we use a 'topologic' approach. If we let $k = 0$, then $x = \alpha t + \beta = x_{\text{in}}$, $y = \mu t + \nu = t$, and (17) takes the form $S(x_{\text{in}}, t) = 0$ or $t = S(x_{\text{in}})$. In that case (19) becomes identical to (3) and we have $\sigma' = \sigma$.

That was for $k = 0$. Now, if we *in a continuous way* vary k from zero to some other value, apparently, t will also be changing in a continuous way. Assuming that $B'^2 - A'C' > 0$, the continuous change of t means that σ' doesn't change it's value. Now, $B'^2 - A'C' = 0$ would mean that the line (8) is a tangent of $y = S(x)$, which cannot happen under given conditions. Therefore, the value of σ' is always the same for all k.

Having found t, it's easy to compute the remaining signals in the structure.

Notice, that if we use the approximation (16) for $S(x)$, then $x_0 = y_0 = 0$ and the segment selection search reduces to checking the sign of β ($\beta \geq 0$ corresponding to the $[0, +\infty]$ segment of $S(x)$).

Now, let's assume $\beta \geq 0$. Since $\alpha < 1$, we have $B' = \beta - \alpha + 1 \geq 0$. Since $\sigma = -1$, we have $\sigma' = -1$. Therefore $\sigma' B' \leq 0$ and we can always use (20b).

Similarly, for the negative segment we get $B' = \beta + \alpha - 1 \leq 0$, $\sigma' = \sigma = 1$. Therefore again $\sigma' B' \leq 0$ and we can always use (20b) as for the positive segment.

7 State-variable filter

Another example which we are going to use here, is a 2-pole state-variable filter with a nonlinearity. In Fig. 5, $S^{-1}(x)$ is the nonlinearity (S^{-1} denotes an inverse function of S, the purpose of this notation will become clear later) and $gx + s_1$ and $gx + s_2$ are the instant responses of the two bilinear-transformed integrators. The instant gain g is defined by the cutoff setting and generally can take any positive value ($g > 0$). R is the parameter controlling the resonance: at $R = 0$ there is no resonance, at $R = 1$ the filter gets into the selfoscillation, at $R = 2$ the poles become real (but this time positive, so that the filter is still unstable) again. To ensure a single intersection between $S^{-1}(x)$ and (8) we assume $(S^{-1})' \geq 1$ and $R < 2$.

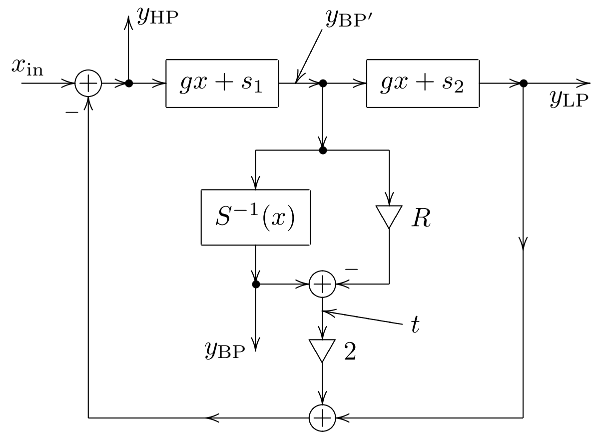

Figure 5: 2-pole state-variable filter with a nonlinearity.

If $S^{-1}(x) \approx x$ for small x, then at small signal levels the filter turns into a classical 2-pole state-variable filter with the corresponding analog prototype transfer function equal to

$$H(s) = \frac{k_{\text{LP}} + k_{\text{BP}} s + k_{\text{HP}} s^2}{1 + 2(1 - R)s + s^2}$$

(written in the normalized-cutoff form $\omega_c = 1$, the variables k_{LP}, k_{BP}, k_{HP} obviously being the output-mixing coefficients).

Now, the purpose of the nonlinearity in this filter is to boost the damping signal if the output level of the first integrator becomes too high. Therefore, rather than being a saturator, $S^{-1}(x)$ needs to be an anti-saturator or 'booster', which can be achieved by e.g. using the areatangent (the inverse hyperbolic tangent) curve $y = \tanh^{-1} x$ or another curve of a similar shape.

If we assume that $S^{-1}(\pm 1) = \pm\infty$ (which is particularly the case for $\tanh^{-1}$), then obviously $|y_{\mathrm{BP}'}| < 1$. To keep the bandpass output level comparable to the low- and high-pass outputs, we let the bandpass signal $y_{\mathrm{BP}'}$ be transformed by the same booster $S^{-1}(x)$, picking up the result y_{BP} as the bandpass output of the filter.[7]

The choice of the unknown t is not so obvious with this filter. Letting t to be picked up at the output of the nonlinear element ($t = y_{\mathrm{BP}}$), as we did in the ladder filter case, is not very good, since if $1 - 2gR + g^2 = 0$ (which occurs e.g. for $g = R = 1$), then it's not possible to determine the remaining signals in the structure from t at all, except by evaluating the booster inverse $S(y_{\mathrm{BP}})$, which we would rather avoid for the computation complexity reasons. Picking t at the nonlinearity's input instead ($t = y_{\mathrm{BP}'}$) has the nice property of keeping the sign of $\sigma' B'$ constant for $S(x)$ defined by (16), but computing y_{HP} and y_{BP} from $y_{\mathrm{BP}'}$ can lead to significant precision losses in certain situations (we might still use it, if we are interested only in the lowpass output signal). Letting $t = y_{\mathrm{HP}}$ is probably okay, however, in this article we choose t as shown in the picture.

Having found t, we may compute y_{HP} as

$$y_{\mathrm{HP}} = \frac{(x - s_2) - gs_1 - 2t}{1 + g^2}$$

and it's better to compute y_{BP} as

$$y_{\mathrm{BP}} = t + Ry_{\mathrm{BP}'}$$

rather than $y_{\mathrm{BP}} = S^{-1}(y_{\mathrm{BP}'})$ since the latter, besides higher computational complexity, has greater precision losses.

For t chosen as described above, the equation (7) becomes:

$$S((1 + R\mu)t + R\nu, \mu t + \nu) = 0 \tag{21}$$

where

$$\begin{aligned} \mu &= -\frac{2g}{1 + g^2} \\ \nu &= \frac{g(x_{\mathrm{in}} - s_2) + s_1}{1 + g^2} \end{aligned}$$

(obviously, $-1 \le \mu < 0$).

Notice that we wrote (21) relative to S rather than S^{-1}. This will allow us to use the saturating approximation curves (such as (16)).

The signed line distances for this filter are

$$d_i = x_i\mu - y_i(1 + R\mu) + \nu$$

(where $y_i = S(x_i)$).

Since $(S^{-1})' \ge 1$, we have $0 < S(x) \le 1$ and (taking into account that $1 + R\mu > \mu$) we conclude that the sequence $\{d_i\}$ is strictly descending. Thus, the index of the active segment can be found by a simple binary search.

Substituting (21) into (1) we obtain

$$A't^2 - 2B't + C' = 0 \tag{22}$$

where

$$\begin{aligned} A' &= P'\mu^2 + 2Q'\mu + a \\ B' &= -P'\mu\nu + S'\mu - Q'\nu + p \\ C' &= P'\nu^2 - 2S'\nu + r \end{aligned} \tag{23}$$

where

$$\begin{aligned} P' &= aR^2 - 2bR + c \\ Q' &= aR - b \\ S' &= pR + q \end{aligned}$$

and therefore

$$\begin{aligned} t &= \frac{B' + \sigma'\sqrt{B'^2 - A'C'}}{A'} \\ &= \frac{C'}{B' - \sigma'\sqrt{B'^2 - A'C'}} \end{aligned} \tag{24}$$

To find out the value of σ' we again use the 'topological' approach. This time we let $R = 1$, $g = 1$ so

[7] Notice, that while for the ladder filter described earlier in the text, the filter output signal didn't exceed $\sup |S(x)|$, the same is no longer applicable for the state-variable filter model we are considering now.

that (22) turns into

$$a\nu^2 - 2b\nu(\nu - t) + c(\nu - t)^2 - 2p\nu - 2q(\nu - t) + r = 0 \quad (25)$$

By letting $x = \nu$ and $y = \nu - t$ we turn (25) into (1) from where

$$y = \nu - t = \frac{B + \sigma\sqrt{B^2 - AC}}{A} \quad (26)$$

Expressing t from (26) and comparing it to (24) we conclude that $\sigma' = -\sigma$.

Using the approximation (16) for $S(x)$ we have $x_0 = y_0 = 0$, which reduces the segment selection rule to checking the sign of ν ($\nu \geq 0$ corresponding to the $[0, +\infty]$ segment of $S(x)$).

Contrarily to the ladder filter case, we don't have a constant sign of B' on either of the segments, therefore it's probably best to use the both forms of (24).

8 Conclusion

We have just described a family of methods allowing analytical computation of a zero-delay feedback loop containing a single nonlinearity. Two examples of application have been given.

While the article has been primarily dealing with piecewise approximations by second-order segments in the most general form (1), more specific forms may be enforced. By forcing $c = 0$ the function spectrum is restricted to rational functions of the form (5). Further forcing $a = 0$ we make the functions take the simple hyperbolic form

$$y = \frac{1}{2} \cdot \frac{-2px + r}{bx + q} \quad (27)$$

Letting $b = 0$ converts (27) into a linear function. Of course, the conditions (12) may need to be adjusted according to the number of degrees of freedom.

In the examples we have considered, the equation (6) was always having a single solution. The cases where multiple solutions are possible can be handled by using the nonzero-impedance approach [1] to pick up the right solution. The segment search algorithms need to be modified as well.

The method can be relatively easily extended to the case of multiple nonlinearities contained in a *single* feedback loop. If we restrict the approximating segments to (27), then a serial chain of multiple nonlinearities interleaved by LTI subsystems will still result in a quadratic feedback equation. The segment search algorithms again require a modification.

Acknowledgements

The author would like to thank NI, and Stephan Schmitt in particular, for their support and for helping him to release this article.

References

[1] Zavalishin V. "Preserving the LTI system topology in *s*- to *z*-plane transforms"