

# New hyperbolic roll-off curves for Darktable's Filmic module

Mr. Teatime (mr.teatime@posteo.de)

March 9, 2021

## Abstract

Until v3.4, Filmic has been using polynomials to model the roll-off of shadows and highlights at the upper and lower end of the dynamic range. This is a mathematical description of hyperbolic functions to replace them. These avoid overshooting, are easier to control and are expected to more realistically reflect the behaviour of photographic film.

## Contents

<b>1</b>	<b>Task</b>	<b>1</b>
<b>2</b>	<b>Adapting a generic hyperbolic curve</b>	<b>2</b>
2.1	Matching the curve to prescribed inputs . . . . .	3
<b>3</b>	<b>Bounds and Constraints</b>	<b>4</b>
3.1	Mathematical requirements . . . . .	4
3.2	Useful range of $b$ . . . . .	5
<b>4</b>	<b>Summary of the construction process</b>	<b>6</b>
4.1	Input Data . . . . .	6
4.2	Curve construction . . . . .	7
4.3	Evaluation . . . . .	7

## 1 Task

Filmic uses a sigmoid curve to convert from scene-referred (linear) colour space with potentially very large dynamic range data to the output colourspace, which is display-referred and thus has a fixed maximum brightness. In the central range of the mapping, the output brightness is proportional to the logarithm of the input brightness. Above and below the linear range, some function is needed to “compress” the brightness information which cannot be represented directly in the output colourspace. The new function needs to:

- continue tangentially from the ends of the linear segment
- end at the maximum/minimum available brightness of the output colourspace, for a user-specified value in the input data

- smoothly reduce the brightness gradient towards the end to approximate the behaviour of photographic film
- not overshoot under any circumstances
- not require numeric operations to obtain the mapping function
- provide a non-zero gradient at the extreme ends to ensure non-zero contrast for all values within the user-specified limits of the mapping

The choice of function in equal parts on Michaelis-Menten functions <sup>1</sup> and the sigmoid functions used in the original publication of the “filmic” tone mapping <sup>2</sup>.

## 2 Adapting a generic hyperbolic curve

The new curves are based on the generic rational function shown in eq. 1

$$\psi(\chi) = a \frac{b\chi^2 + \chi}{b\chi^2 + \chi + c} \quad (1)$$

This has three coefficients  $a$ ,  $b$  and  $c$ , and produces an asymptotic curve, which goes through  $(0,0)$  and asymptotically approaches  $a$  as  $\chi \rightarrow \infty$ , as shown in figure 2.

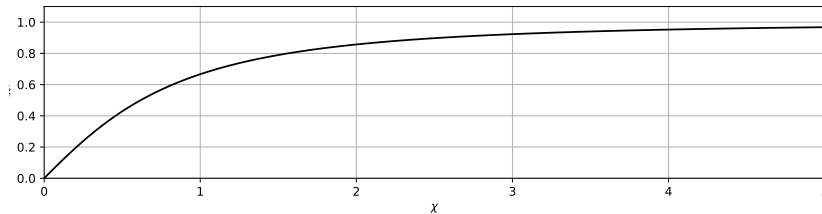


Figure 1: Curve resulting from equation 1, with all parameters set to 1.

With the available coefficients we can change the asymptote, the initial tangent, and have one degree of freedom left to influence the curve shape (to some extent). The first derivative of our curve is (derivation in appendix):

$$\psi'(\chi) = ac \frac{2b\chi + 1}{(b\chi^2 + \chi + c)^2} \quad (2)$$

The second derivative is rather more complicated and is only shown for  $\chi = 0$  here (derivation in appendix):

$$\psi''(\chi = 0) = 2a \left( \frac{b}{c} - \frac{1}{c^2} \right) \quad (3)$$

<sup>1</sup>[https://en.wikipedia.org/wiki/Michaelis-Menten\\_kinetics](https://en.wikipedia.org/wiki/Michaelis-Menten_kinetics)

<sup>2</sup>See slides 55 and 56 in this presentation from Haarm-Pieter Duiker: <http://duikerresearch.com/2015/09/filmic-tonemapping-for-real-time-rendering>

## 2.1 Matching the curve to prescribed inputs

The curve in filmic needs to start at a given point with a given tangent, and end at another given point. To simplify the derivation and calculation of coefficients, the problem is solved in two steps:

1. Derive the solution for a curve passing through the origin, with the second point in the positive quadrant, in the “local” construction space defined by  $\chi$  and  $\psi$
2. “move” the resulting curve to its desired position by substituting:

$$\chi = |x - x_0| \quad (4)$$

$$\psi(\chi) = |y - y_0| \quad (5)$$

The equations to match the curve to the prescribed tangend  $g$  and the end point  $(\chi_1; \psi_1)$  solve are:

$$\psi'(\chi = 0) = g = ac \frac{2b\chi + 1}{(b\chi^2 + \chi + c)^2} \Big|_{\chi=0} \quad (6)$$

$$\psi'(\chi = \chi_1) = \psi_1 = a \frac{b\chi_1^2 + \chi_1}{b\chi_1^2 + \chi_1 + c} \quad (7)$$

Equation 6 evaluates to:

$$g = \frac{a}{c} \Rightarrow a = gc \quad (8)$$

Inserting equation 8 into equation 7 yields:

$$\begin{aligned} gc \frac{b\chi_1^2 + \chi_1}{b\chi_1^2 + \chi_1 + c} &= \psi_1 \\ gc (b\chi_1^2 + \chi_1) &= \psi_1 (b\chi_1^2 + \chi_1 + c) \\ gc (b\chi_1^2 + \chi_1) - \psi_1 c &= \psi_1 (b\chi_1^2 + \chi_1) \\ c \left( b\chi_1^2 + \chi_1 - \frac{\psi_1}{g} \right) &= \frac{\psi_1}{g} (b\chi_1^2 + \chi_1) \\ c &= \frac{\psi_1}{g} \frac{b\chi_1^2 + \chi_1}{b\chi_1^2 + \chi_1 - \frac{\psi_1}{g}} \end{aligned} \quad (9)$$

This means that the boundary conditions can be fulfilled using just the two coefficients  $a$  and  $c$ , which means that  $b$  is available for use a user-definable free curve parameter. The range within which this is possible without preventing issues with solving the equations is addressed in section 3.

To produce the highlight roll-off curve which originates at  $(x_{0hi}, y_{0hi})$  with the tangent  $g$  and ends at  $(x_{max}, y_{max})$ , the substitution yields:

$$\chi_{hi} = x - x_{0hi}, \quad x \in [x_{0hi}; x_{max}] \quad (10)$$

$$\chi_{1hi} = x_{min} - x_{0hi} \quad (11)$$

$$c_{hi} = \frac{y_{max} - y_{0hi}}{g} \frac{b\chi_{1hi}^2 + \chi_{1hi}}{b\chi_{1hi}^2 + \chi_{1hi} - \frac{y_{max} - y_{0hi}}{g}} \quad (12)$$

$$a_{hi} = c_{hi}g \quad (13)$$

$$y_{hi}(x) = y_{0hi} + a_{hi} \frac{b_{hi}\chi_{hi}^2 + \chi_{hi}}{b_{hi}\chi_{hi}^2 + \chi_{hi} + c_{hi}} \quad (14)$$

For the shadow range, the curve needs to match  $(x_{0low}, y_{0low})$  (with the same tangent  $g$  as the highlight curve), and end at  $(x_{min}, y_{min})$ :

$$\chi_{lo} = x_{0lo} - x, \quad x \in [x_{0lo}; x_{max}] \quad (15)$$

$$\chi_{1lo} = x_{0lo} - x_{min} \quad (16)$$

$$c_{lo} = \frac{y_{max} - y_{0lo}}{g} \frac{b\chi_{1lo}^2 + \chi_{1lo}}{b\chi_{1lo}^2 + \chi_{1lo} - \frac{y_{max} - y_{0lo}}{g}} \quad (17)$$

$$a_{lo} = c_{lo}g \quad (18)$$

$$y_{lo}(x) = y_{0lo} - a_{lo} \frac{b_{lo}\chi_{lo}^2 + \chi_{lo}}{b_{lo}\chi_{lo}^2 + \chi_{lo} + c_{lo}} \quad (19)$$

### 3 Bounds and Constraints

#### 3.1 Mathematical requirements

The “useful” range of the curve (within which it provides the desired behaviour of monotonic increase requires a positive first derivative, which according to equation 2 means:

$$ac \frac{2b\chi + 1}{(b\chi^2 + \chi + c)^2} > 0 \quad (20)$$

$$\Rightarrow ac(2b\chi + 1) > 0 \quad (21)$$

Since the equation has been set up to only be evaluated at  $\chi \geq 0$ , this means that  $b$  needs to be positive to avoid the right-hand term from becoming 0, which means that  $ac > 0$ . Since  $g$  is always positive in the application at hand,  $a$  and  $c$  always have the same sign, since they are linked by equation 8. This means that  $c$  must be greater than zero to provide a monotonic curve. According to equation 9 and using positive input for  $\psi_1$ , this is the case if:

$$b\chi_1^2 + \chi_1 - \frac{\psi_1}{g} > 0 \quad (22)$$

$$\Rightarrow b\chi_1^2 + \chi_1 > \frac{\psi_1}{g} \quad (23)$$

$$\Rightarrow g(b\chi_1 + 1) > \frac{\psi_1}{\chi_1} \quad (24)$$

Since  $b \geq 0$ , this means that in the worst case (of  $b = 0$ ),  $g$  must be greater than  $\frac{\psi_1}{\chi_1}$ . Geometrically, that corresponds to the slope of a straight line connecting  $(0, 0)$  to  $(\chi_1, \psi_1)$ , and is the lowest slope which can be useful in the context of tonemapping, as otherwise the “roll-off”-curve would need to exceed the slope of the central linear segment.

Regarding the computation of coefficients,

$$\begin{aligned} \chi_1 \neq 0 & \quad ; \quad \psi_1 \neq 0 \\ \Rightarrow x_1 \neq x_0 & \quad ; \quad y_1 \neq y_0 \end{aligned} \quad (25)$$

The gradient  $g$

### 3.2 Useful range of $b$

The curve parameter  $b$  affects the shape of the roll-off curve and has an effect on the second derivative at the hand-over points to the linear segment. Equation 3 shows that  $b$  can affect the sign of the second derivative at the hand-over point, which should usually never become positive (a positive second derivative corresponds to the curve “curving up”). As shown in the previous section,  $b$  also must be positive (but may be zero).

To determine the maximum value which  $b$  can take without changing the sign of the second derivative – which simultaneously is also the value at which the second derivative becomes zero at the hand-over point – we set the equation 3 to zero:

$$\begin{aligned}\psi''(\chi = 0, b = b_{max}) &= 0 = 2a \left( \frac{b_{max}}{c} - \frac{1}{c^2} \right) \\ \Rightarrow b_{max} &= \frac{1}{c}\end{aligned}\tag{26}$$

Inserting the general solution for  $c$  from equation 9:

$$\begin{aligned}b_{max} &= \frac{g}{\psi_1} \frac{b_{max}\chi_1^2 + \chi_1 - \frac{\psi_1}{g}}{b_{max}\chi_1^2 + \chi_1} \\ b_{max} (b_{max}\chi_1^2 + \chi_1) &= \frac{g}{\psi_1} \left( b_{max}\chi_1^2 + \chi_1 - \frac{\psi_1}{g} \right) \\ b_{max}^2\chi_1^2 + b_{max}\chi_1 - \frac{g}{\psi_1}b_{max}\chi_1^2 &= g\frac{\chi_1}{\psi_1} - 1 \\ b_{max}^2\chi_1^2 + b_{max} \left( \chi_1 - g\frac{\chi_1^2}{\psi_1} \right) &= g\frac{\chi_1}{\psi_1} - 1 \\ b_{max}^2\chi_1^2 + b_{max} \left( \chi_1 - g\frac{\chi_1^2}{\psi_1} \right) + \frac{1}{4} \left( 1 - g\frac{\chi_1}{\psi_1} \right)^2 &= g\frac{\chi_1}{\psi_1} - 1 + \frac{1}{4} \left( 1 - g\frac{\chi_1}{\psi_1} \right)^2\end{aligned}$$

And applying the first binomial equation ( $(A + B)^2 = A^2 + 2AB + B^2$ ) to the left-hand side and the second binomial equation ( $(A - B)^2 = A^2 - 2AB + B^2$ ) on the right side:

$$\begin{aligned}\left( b_{max}\chi_1 + \frac{1}{2} \left( 1 - g\frac{\chi_1}{\psi_1} \right) \right)^2 &= g\frac{\chi_1}{\psi_1} - 1 + \frac{1}{4} \left( 1 - 2g\frac{\chi_1}{\psi_1} + g^2\frac{\chi_1^2}{\psi_1^2} \right) \\ \left( b_{max}\chi_1 + \frac{1}{2} \left( 1 - g\frac{\chi_1}{\psi_1} \right) \right)^2 &= \frac{1}{4} \left( g^2\frac{\chi_1^2}{\psi_1^2} + 2g\frac{\chi_1}{\psi_1} - 3 \right) \\ b_{max}\chi_1 + \frac{1}{2} \left( 1 - g\frac{\chi_1}{\psi_1} \right) &= \frac{1}{2} \sqrt{g^2\frac{\chi_1^2}{\psi_1^2} + 2g\frac{\chi_1}{\psi_1} - 3} \\ b_{max}\chi_1 &= \frac{1}{2} \sqrt{g^2\frac{\chi_1^2}{\psi_1^2} + 2g\frac{\chi_1}{\psi_1} - 3} + \frac{g}{2} \frac{\chi_1}{\psi_1} - \frac{1}{2} \\ b_{max}\chi_1 &= \frac{1}{2} \left( g\frac{\chi_1}{\psi_1} \pm \sqrt{g^2\frac{\chi_1^2}{\psi_1^2} + 2g\frac{\chi_1}{\psi_1} - 3 - 1} \right) \\ b_{max} &= \frac{1}{2\chi_1} \left( g\frac{\chi_1}{\psi_1} \pm \sqrt{g^2\frac{\chi_1^2}{\psi_1^2} + 2g\frac{\chi_1}{\psi_1} - 3 - 1} \right)\end{aligned}\tag{27}$$

To analyse the solvability and the signs of the equation, we substitute  $h = g \frac{\chi_1}{\psi_1}$ . Note that due to equation 24,  $h > 1$ :

$$b_{max1} = \frac{1}{2\chi_1} \left( h + \sqrt{h^2 + 2h - 3} - 1 \right) \quad (28)$$

$$b_{max2} = \frac{1}{2\chi_1} \left( h - 1 - \sqrt{h^2 + 2h - 3} \right) \quad (29)$$

The term within the square root is always positive, which guarantees that  $b_{max}$  can always be computed.  $h - 1$  is also always positive, which guarantees that  $b_{max1}$  is always positive.  $b_{max2}$ , on the other hand, is never larger than zero:

$$\begin{aligned} b_{max2} &= \frac{1}{2\chi_1} \left( h - 1 - \sqrt{h^2 + 2h - 3} \right) \leq 0 \\ \Rightarrow & \quad \quad \quad h - 1 \leq \sqrt{h^2 + 2h - 3} \\ \Rightarrow & \quad \quad \quad \sqrt[+]{h^2 - 2h + 1} \leq \sqrt[+]{h^2 + 2h - 3} \\ \Rightarrow & \quad \quad \quad h^2 - 2h + 1 \leq h^2 + 2h - 3 \\ \Rightarrow & \quad \quad \quad 4 \leq 4h \quad \quad \quad q.e.d \end{aligned}$$

This means that although there are two mathematically possible solutions, there is always exactly one positive solution,  $b_{max1}$ .

## 4 Summary of the construction process

This is a more concise summary of the equations needed to implement the hyperbolic parametric curves, including the conditions which need to be met to ensure that the function can always be evaluated and always provides the desired behaviour.

### 4.1 Input Data

The function represents a transition from  $(x_0, y_0)$  to  $(x_1, y_1)$ , with a gradient of  $g$  imposed at the starting point. Following equation 25, the points may be in any quadrant, as long as both points do not have neither the same  $x$  nor  $y$  coordinate. The gradient  $g$  needs to have the same sign as that of a straight line between the two input points, but be steeper than this line (see equation 24):

$$g = m \frac{y_1 - y_0}{x_1 - x_0} \quad , \quad m > 1 \quad (30)$$

Please note that  $m = 1$  is not included in the valid range, but  $m = 1 + \epsilon$  (using some small multiple of machine epsilon) is safe. The valid range can be made to include  $m = 1$  if  $m$  is computed on user input. If  $m = 1$ , the roll-off curve then needs to be replaced with a straight line function, since that is what the hyperbolic function approaches for  $m = 1$  (but cannot represent since it requires a horizontal asymptote, in this case at infinity).

The limits on the curve parameter  $b$  strongly depends on the inputs, and needs to be between 0 and informed by equation 28:

$$b = \frac{\beta}{2\chi_1} \left( g \frac{\chi_1}{\psi_1} - 1 + \sqrt{g^2 \frac{\chi_1^2}{\psi_1^2} + 2g \frac{\chi_1}{\psi_1} - 3} \right) , \quad \beta \in [0; 1] \quad (31)$$

## 4.2 Curve construction

The input points are transformed to construction space by translating them such that the starting point moves to  $(0; 0)$ , the relative distance between both points is unchanged, but the end point is always in the first quadrant:

$$\begin{aligned} \chi_1 &= |x_1 - x_0| \\ \psi_1 &= |y_1 - y_0| \end{aligned} \quad (32)$$

The sign of the gradient would also need to be adapted to be positive to compute the curve parameters, but within *filmic*,  $g$  is already always positive. This would only change if Darktable or *filmic* were to provide the option of a negative mapping at some point in the future. This could be dealt with swiftly by simply always discarding the sign and using the absolute value of  $g$  to compute the curve coefficients.

The curve coefficients are then computed using equations 9 and 8

## 4.3 Evaluation

Transforming sample  $x$  coordinates is done using equation 32:

$$\chi(x) = |x - x_0| \quad (33)$$

the resulting  $y$  as well as local gradient can be computed as follows, using  $\psi(\chi)$  (equation 1) and its derivative (equation 2):

$$y(x) = y_0 + \text{sgn}(y - y_0) \psi(\chi(x)) \quad (34)$$

$$y'(x) = \text{sgn}(g) \psi'(\chi(x)) \quad (35)$$

## Appendix

baseline equation (eq. 1):

$$\psi(\chi) = a \frac{b\chi^2 + \chi}{b\chi^2 + \chi + c}$$

derived, using quotient rule:

$$\begin{aligned} \frac{d\psi(\chi)}{d\chi} &= a \frac{(2b\chi + 1)(b\chi^2 + \chi + c) - (b\chi^2 + \chi)(2b\chi + 1)}{(b\chi^2 + \chi + c)^2} \\ &= a \frac{(2b\chi^3 + b\chi^2 + 2b\chi^2 + \chi + 2bc\chi + c) - (2b\chi^3 + b\chi^2 + 2b\chi^2 + \chi)}{(b\chi^2 + \chi + c)^2} \\ &= a \frac{2bc\chi + c}{(b\chi^2 + \chi + c)^2} \\ \psi'(\chi) &= ac \frac{2b\chi + 1}{(b\chi^2 + \chi + c)^2} \end{aligned}$$

This is equation 2.

To compute the second derivative, we first expand the denominator and compute its derivative:

$$(b\chi^2 + \chi + c)^2 = b^2\chi^4 + 2b\chi^3 + (2bc + 1)\chi^2 + 2c\chi + c^2 \quad (36)$$

$$\frac{d(b^2\chi^4 + 2b\chi^3 + (2bc + 1)\chi^2 + 2c\chi + c^2)}{d\chi} = 4b^2\chi^3 + 6b\chi^2 + (4bc + 2)\chi + 2c \quad (37)$$

With equations 36 and 37, we can then apply the quotient rule to the first derivative equation:

$$\begin{aligned} \frac{d\psi^2(\chi)}{d\chi^2} &= ac \\ \frac{2b(b^2\chi^4 + 2b\chi^3 + (2bc + 1)\chi^2 + 2c\chi + c^2) - (2b\chi + 1)(4b^2\chi^3 + 6b\chi^2 + (4bc + 2)\chi + 2c)}{(b\chi^2 + \chi + c)^4} \end{aligned} \quad (38)$$

Since the second derivative is only used for  $\chi = 0$ , the expression in equation 38 is not further simplified but only the expression for  $\chi = 0$  is listed here:

$$\begin{aligned} \psi''(\chi = 0) &= ac \frac{2b(0 + c^2) - (0 + 2c)}{(0 + c)^4} \\ &= ac \frac{2bc^2 - 2c}{c^4} \\ &= 2a \frac{bc - 1}{c^2} \\ &= 2a \left( \frac{b}{c} - \frac{1}{c^2} \right) \end{aligned}$$