Bivariate Loss Functions:

a general method for measuring accuracy of approximation

Andy Allinger

Kruger Optical Corporation

6 November 2022

Abstract: A novel class of loss functions is introduced for the estimation problem where training data is available. The proposed functions depend on the target value and the prediction separately, subject to optimality and convexity constraints. Necessary and sufficient conditions are specified. Many examples are illustrated. The proposed class has potential for application in function approximation and machine learning.

Consider the function interpolation problem of a real-valued scalar function of several real variables, in which training data of *m* instances is available in the form ($\mathbf{x}_{ij}$, $\mathbf{y}_i$), where $\mathbf{y}_i$ is the thing to be predicted, and $\mathbf{x}_{ij}$ is the *n*-dimensional data available on which to make the prediction. An interpolation is sought of the form $\hat{y} = f(x_{ij}, \theta_k)$ where *f* may be any function of $\mathbf{x}_{ij}$ and of the *p* parameters $\boldsymbol{\theta}_k$. This problem may be approached by casting it as an optimization problem of the form:

$$\theta \;=\; \operatorname{argmin} J(y, f(x, \theta))$$

where J is some suitable loss function. For sake of clarity, let *z* denote the estimate $\hat{y}$. Then it is possible to write J(**y**, **z**) for the function to be minimized. Often this is the least squares loss:

$$J \;=\; \sum_{i=1}^{m} (z_i - y_i)^2$$

or the least absolute values (LAV) loss:

$$J \;=\; \sum_{i=1}^{m} |z_i - y_i|$$

which has been in use since Galileo [Bidabad, 2005]. This article proposes a new class of loss functions in which the loss per training example is not necessarily a function of the difference between the prediction and the target value *z-y*, but may be a general function *g(z,y)*, subject to certain restrictions described in the next section. Continue to assume that the loss is the sum of the separate penalties of each instance of the training data, such that:

$$J = \sum_{i=1}^{m} g(y_i, z_i)$$

Within this structure a broad class of functions may be considered for *g(y, z)*, which may prove more suitable in certain applications than the least squares or least absolute values objectives.

The terms loss function, cost function, and objective function are used interchangeably in this article, care only being required to distinguish the overall loss *J* summed over the training set from the component loss *g* of a particular instance.

Criteria

Consider the necessary conditions which must be fulfilled by any objective function *g(y,z)*. First, require *g* to be continuous and twice differentiable in *z*.

Optimality. A necessary condition is

$$\left.\frac{\partial g}{\partial z}\right|_{z=y} = 0 \qquad \text{(i)}$$

Convexity. For optimization by gradient descent, the conditions are necessary

$$\left.\frac{\partial g}{\partial z}\right|_{z<y} < 0 \qquad \text{(ii)}$$

$$\left.\frac{\partial g}{\partial z}\right|_{z>y} > 0 \qquad \text{(iii)}$$

Definiteness. Impose the condition that the penalty is zero for a correct prediction:

$$\left. g\right|_{z=y} = 0 \qquad \text{(iv)}$$

As a further condition, this article will consider only the case in which

$$\frac{\partial^2 g}{\partial z^2} > 0, \quad \forall z \qquad \text{(v)}$$

which given (i) is sufficient but not necessary to ensure (ii) and (iii). Thus, for any strictly positive expression in *z*, a related loss function may be obtained by integration.

Notice that for the least squares and least absolute values component functions, any constant factor *u* introduced such that $g = (uz - uy)^2$ may be factored out of the overall objective $J = u^2 \sum_i (z_i - y_i)^2$ meaning that the preference between machine learning models is independent of the unit of measure of *y*, so that

$$J(z_1 = f_1(x, \theta_1), y) < J(z_2 = f_2(x, \theta_2), y) \;\Leftrightarrow\; u^2 J(z_1 = f_1(x, \theta_1), y) < u^2 J(z_2 = f_2(z, \theta_2), y)$$

Symmetric Loss Functions

Consider first a simple example. Let *p* denote a parameter measured in the same units as *z*, and let b denote a positive constant measured in units of loss per units of *z*. Choose as a strictly positive expression:

$$\frac{\partial^2 g}{\partial z^2} = \frac{b}{p^2} > 0$$

Integrate:

$$\frac{\partial g}{\partial z} = \frac{b}{p^2} \int dz = \frac{b}{p^2}(z + c_1)$$

Apply condition (i):

$$\left.\frac{\partial g}{\partial z}\right|_{z=y} = 0 = \frac{b}{p^2}(y + c_1)$$
$$c_1 = -y$$
$$\frac{\partial g}{\partial z} = \frac{b}{p^2}(z - y)$$

Integrate:

$$g = \frac{b}{p^2} \int (z - y)\, dz = \frac{b}{p^2}\left(\frac{1}{2} z^2 - yz + c_2\right)$$

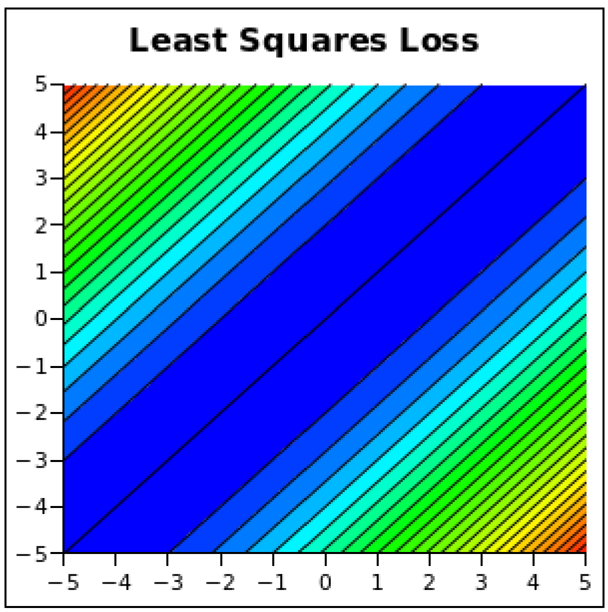
Apply condition (iv):

$$g|_{z=y} = 0 = \frac{b}{p^2}\left(\frac{1}{2} y^2 - y^2 + c_2\right)$$
$$c_2 = \frac{1}{2} y^2$$
$$g = \frac{b}{2}\left(\frac{z - y}{p}\right)^2$$

Setting $b = p = 1$ and neglecting the constant factor of ½ gives

$$g = (z - y)^2$$

and so the least squares loss proves to be an especially simple case of the broader class of bivariate loss functions.



Contours of least-squares loss are parallel and symmetric.

As another familiar example, take the step function:

$$\frac{\partial g}{\partial z} = \begin{Bmatrix} -1, & z<y \\ \text{undefined}, & z=y \\ +1, & z>y \end{Bmatrix}$$

which satisfies (ii) and (iii) and does not violate (i) in the sense that $\frac{\partial g}{\partial z}$ is undefined at zero. Thus,

$$g = \int \begin{Bmatrix} -1, & z<y \\ +1, & z>y \end{Bmatrix} dz = \begin{Bmatrix} -z + c_n, & z<y \\ +z + c_p, & z>y \end{Bmatrix}$$

Apply condition (iv)

$$g\big|_{z=y} = 0 = -y + c_n = +y + c_p$$
$$c_n = y$$
$$c_p = -y$$

So

$$g = \begin{Bmatrix} -z + y, & z<y \\ z - y, & z>y \end{Bmatrix} = |z-y|$$

which admittedly would not have been discovered under the present criteria but can be seen to be consistent with them.

Asymmetric Loss Functions

The LINEX (LINear EXponential) function, attributed to Hal R. Varian, 1975, is given by

$$L(\Delta) = b\mathrm{e}^{a\Delta} - c\,\Delta - b$$

It has the interesting property of merging a linearly increasing penalty for underestimation with a exponentially increasing penalty for overestimation (or vice versa depending on the parameters.) [Zellner, 1986] observes "for a minimum to exist at Δ = 0, we must have *ab* = *c*" and thus:

$$L(\Delta) = b\left(\mathrm{e}^{a\Delta} - a\Delta - 1\right)$$

To derive the LINEX function, begin with an expression which satisfies condition (v) for $a \neq 0$.

$$\frac{\partial^2 g}{\partial z^2} = a^2 \mathrm{e}^{(a(z-y))} > 0$$
$$\frac{\partial g}{\partial z} = a^2 \int \mathrm{e}^{(a(z-y))} dz = a\,\mathrm{e}^{(a(z-y))} + c_1$$

Apply condition (i)

$$\left.\frac{\partial g}{\partial z}\right|_{z=y} = 0 = a\, e^{(a(y-y))} + c_1$$

$$c_1 = -a$$

$$\frac{\partial g}{\partial z} = a\, e^{(a(z-y))} - a$$

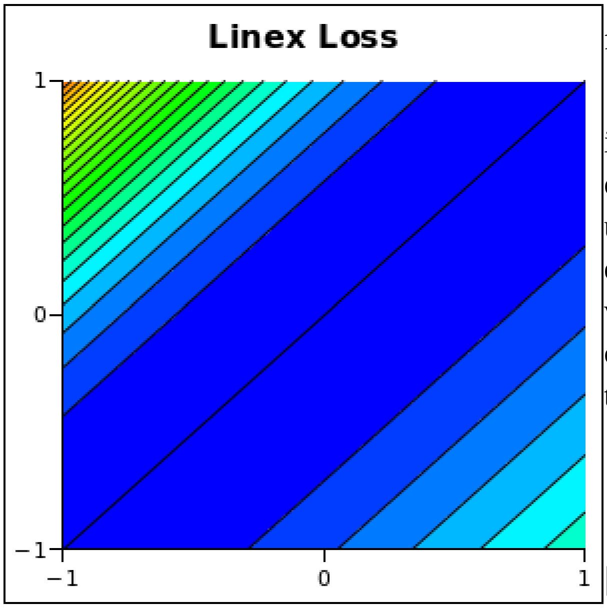
$$g = a\int \left(e^{(a(z-y))} - 1\right) dz = e^{(a(z-y))} - az + c_2$$

Apply condition (iv)

$$g|_{z=y} = 0 = e^{(a(y-y))} - ay + c_2$$

$$c_2 = ay - 1$$

$$g = e^{(a(z-y))} - a(z-y) - 1$$



Contours of LINEX loss are parallel but not symmetric.

which may be recognized as Varian's LINEX loss function for $b = 1$, where $\Delta = z - y$.

The introduction of asymmetric loss is an important development for applications where overestimation may be more costly than underestimation, or vice versa. For example, in dam design [Zellner, 1986] overestimation of peak water levels may cause excessive construction costs, but underestimation may cause the breach of the dam.

Bivariate Loss Functions

The loss functions considered so far have all been functions of the difference $z - y$. It may be advantageous in some circumstances to allow the more general case in which $g(y,z)$ can not be reduced to a function of the single variable $\Delta = z - y$. A bivariate loss function depends on both the training data and the prediction. To motivate the concept, consider a machine-learning system that predicts winners in a horse race. The situation is asymmetric. There are many races on which to bet. Underestimating a horse's speed is a missed opportunity, but there are many such opportunities. Overestimating a horse's speed, on the other hand, leads to the irretrievable loss of money. The function therefore must penalize overestimation more harshly than underestimation, but not in a constant manner across all instances. Since the gambler will not bet horses he predicts will lose, these cases are unimportant. He must be able to distinguish good horses from bad horses, but it is needless for him to distinguish bad horses from terrible horses.

As a typical example, consider an agronomist concerned with improving a breed of hog. Based on the curliness of its tail at birth, he would like to predict its weight at slaughter. The objective is asymmetric. Overestimation is more harmful than underestimation, because meager research funding affords only a few hogs to be raised, while piglets are cheaply available.

It is not required for the weights of small hogs to be evaluated very precisely. They will not be selected, and are in a sense irrelevant. Attention must be given to examples with large positive values. Take the expression satisfying (v):

$$\frac{\partial^2 g}{\partial z^2} = \left((z-y)^2+1\right) e^z > 0$$

Nondimensionalize the expression by introducing parameters *a* and *b* such that:

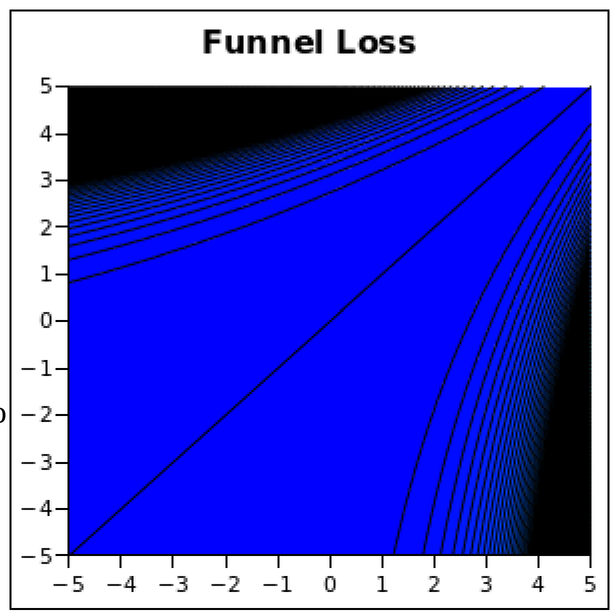
$$\frac{\partial^2 g}{\partial z^2} = \frac{b^2}{a^4}\left(\left(\frac{z-y}{b}\right)^2+1\right) e^{z/a}$$

Then by aid of computer algebra:

$$\frac{\partial g}{\partial z} = \frac{1}{a}\left\{\left[\left(\frac{z-y}{a}\right)^2 - 2\left(\frac{z-y}{a}\right) + \frac{b^2}{a^2} + 2\right] e^{z/a} - \left[\frac{b^2}{a^2}+2\right] e^{y/a}\right\}$$

$$g = \left[\left(\frac{z-y}{a}\right)^2 - 4\frac{z-y}{a} + \frac{b^2}{a^2} + 6\right] e^{z/a} - \left[\left(\frac{b^2}{a^2}+2\right)\frac{z-y}{a} + \frac{b^2}{a^2} + 6\right] e^{y/a}$$

Let this component function be called the "Funnel" for the shape of its graph. The contours are not parallel, because the loss is not a function of the difference *z-y*. Loss increases with increasing error more quickly for positive values than for negative values. Thus, the machine-learning system does not need to adapt the parameters as much for the small instances.



Contours of funnel loss are neither parallel nor symmetric.

As a further ludicrous example, consider the case of a delivery boy who must bring pizza to a haunted house while avoiding a malevolent ghost. The ghost may appear at any time, but can only do injury to mortals around midnight. The delivery boy, being of a scientific turn of mind, has collected the schedules of former delivery boys who have perished under mysterious circumstances, as well as suitable data on zodiac signs, water level in the Nile River, and pronouncements of the Federal Reserve chairmen, and wishes to use this information to optimize the parameters of a machine learning system for predicting the time of appearance of the ghost.

To avoid encountering the ghost, occurrences with a small $|y|$ must be estimated precisely, while large $|y|$ may be roughly approximated. To optimize the parameters of the prediction system, instances (x_{ij},y_i) of small $|y|$ must be emphasized. This can of course be accomplished by weighting the objective function, for example

$$J = \sum_i w_i (z_i - y_i)^2$$

Nevertheless, a loss function that naturally accounts for this is an intriguing possibility worth exploring. Take the threat posed by the ghost and the cost of a missed delivery to be both inversely proportional to the square of the time from away from midnight.[1] However,

$$h = \frac{1}{y^2\ z^2}$$

is not a loss function. Differentiate:

$$\frac{\partial h}{\partial z} = \frac{-2}{y^2}\ z^{-3}$$
$$\frac{\partial^2 h}{\partial z^2} = \frac{6}{y^2}\ z^{-4}$$

This motivates the strictly positive expression:

$$\frac{\partial^2 g}{\partial z^2} = \frac{1}{(y^2 + 1)\ (z^2 + 1)^2} > 0$$

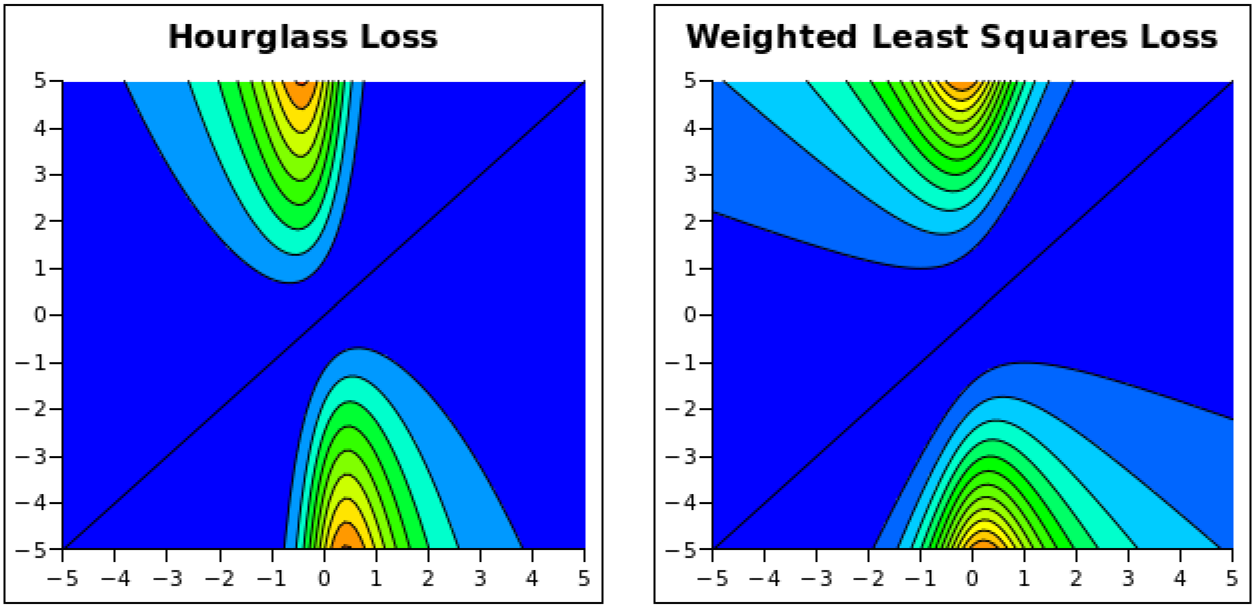
where padding has been added to remove the poles at y=0 and z=0. Introduce scale parameters p and q, then by aid of computer algebra:

$$\frac{\partial^2 g}{\partial z^2} = \frac{1}{q^2}\ \frac{1}{(\frac{y}{p})^2 + 1}\ \frac{1}{\left((\frac{z}{q})^2 + 1\right)^2}$$

$$\frac{\partial g}{\partial z} = \frac{1}{2q}\ \frac{1}{\left(\frac{y}{p}\right)^2 + 1}\left[\arctan\left(\frac{z}{q}\right) - \arctan\left(\frac{y}{q}\right) + \frac{\frac{z}{q}}{\left(\frac{z}{q}\right)^2 + 1} - \frac{\frac{y}{q}}{\left(\frac{y}{q}\right)^2 + 1}\right]$$

$$g = \frac{1}{2}\left(\frac{1}{(\frac{y}{p})^2 + 1}\right)\left(\frac{z}{q}\left(\arctan\left(\frac{z}{q}\right) - \arctan\left(\frac{y}{q}\right)\right) - \frac{z-y}{q}\ \frac{\frac{y}{q}}{(\frac{y}{q})^2 + 1}\right)$$

1 More realistic assumptions would suppose the delivery boy adjusts his arrival to be either earlier or later than the predicted appearance of the ghost. This would result in a discontinuity, putting the problem beyond the reach of ordinary calculus.



Let this component function be called the "Hourglass." It is instructive at this point to compare the contour plots of a weighted least squares component function that downweights training instances far from zero, for example

$$g = \frac{1}{y^2 + 1}(z - y)^2$$

with the hourglass function just derived. Both schemes assign severe penalties for small *y* only. Notice that the plots are not symmetric about the correctness line $y = z$. The cost of estimating that the ghost will appear when it will not is not the same as the cost of estimating that the ghost will not appear when it will.

The present example demands precision for values near zero. The prior example demands precision for large positive values. A third natural possibility is to demand precision for both large positive and large negative values. One such function is given in the appendix.

Discussion

By the linearity of the integral operator, any loss function derived from an expression in *y* results in a weighted least squares objective.

It is interesting to observe that the second integral of a valid objective component function is another objective component function. For example, the second integral of the least-squares loss is the least-fourths loss:

$$g = (z - y)^4$$

Convergence may be slower, because a function for which $\frac{\partial^2 g}{\partial z^2} = 0$ at $z = y$ is nearly flat at its minimum.

All component loss functions have the same minimum, which is zero when the estimate is equal to the training value. They behave differently when summed over a training set. The machine learning system does not reach perfect prediction. Various loss functions lead to various compromises based on the assumptions about which errors are most harmful.

A component function that depends on y can represent the importances of the various events to be predicted. A component function that depends on z can represent the costs of the various actions to be taken in response to the prediction.

Much attention has been given in the literature to alternative loss functions, especially the least absolute values (LAV) or ℓ^1 loss $g = |z - y|$. The LAV loss is a distance between a vector of predictions and training examples under the ℓ^1 metric. The ℓ^1 metric is an instance of the Minkowski ℓ^p metric which proposes that the distance between points in $\mathbb{R}^M$ may be measured as

$$d = \sqrt[p]{\sum_{i=1}^{M} (z_i - y_i)^p}$$

Since low-order ($p < 2$) loss functions are of special interest, it is worth observing that condition (v) eliminates functions that increase more slowly than the absolute value. For example

$$g = \ln\left(1 + (z - y)^2\right)$$

is a convex loss function, but its second derivative

$$\frac{-4(z-y)^2}{\left(1 + (z-y)^2\right)^2} + \frac{2}{1 + (z-y)^2}$$

is negative for $(z - y)^2 > 1$.

A common argument is that LAV is advantageous because it is less sensitive to large values of $z - y$, and that such outliers are likely spurious. In a contrary situation, when the data are quite reliable, it may be preferable to minimize the worst-case error, in which a high-order metric such as ℓ^∞ is appropriate. However, taking into account the costs incurred by different values of y or z, it is by no means certain that the greatest difference $z - y$ is really the worst case.

Another frequently used error measure is the relative error $(z-y)/y$. The present article has restricted attention to functions suitable for all real y and z, and has avoided relative error due to the need to avoid division by zero. A few functions suitable only for positive y and z are given in the appendix. The bottleneck-shaped functions (arctangent, hyperbolic tangent, and normal losses) serve a similar purpose in penalizing heavily the errors on instances of small magnitude.

Conclusion

Bavariate loss functions represent a significant advance towards greater generality. Severity of loss may vary across the range of target values, and may also vary across the range of predictions. This added degree of flexibility allows models to be optimized directly to tasks, rather than optimized for prediction as such. Their application remains to be explored.

References

"Bayesian Estimation and Prediction Using Asymmetric Loss Functions," Arnold Zellner, *J. Am. Statistical Assoc.* v.81 n.394 pp.446-551 Jun 1986. https://www.jstor.org/stable/2289234

"L1 Norm Based Computational Algorithms," Bijan Bidabad, 2005. http://www.bidabad.ir/doc/l1-article6.pdf

A Gallery of Loss Functions

Attached for reference is a listing of loss component functions *g(y,z)* where the overall loss function

$$J = \sum_{i=1}^{m} g(y, z)$$

is a sum of the loss components *g* for each instance over a data set of size *m*.
Let *y* denote the given value to be estimated, and let $z = \hat{y}$ denote the estimate.
Let *a*, *b*, *p*, and *q* denote parameters measured in the same units as *y* and *z*.
Let α denote a parameter measured inverse to the units of *y* and *z*.

Functions for all real y and z

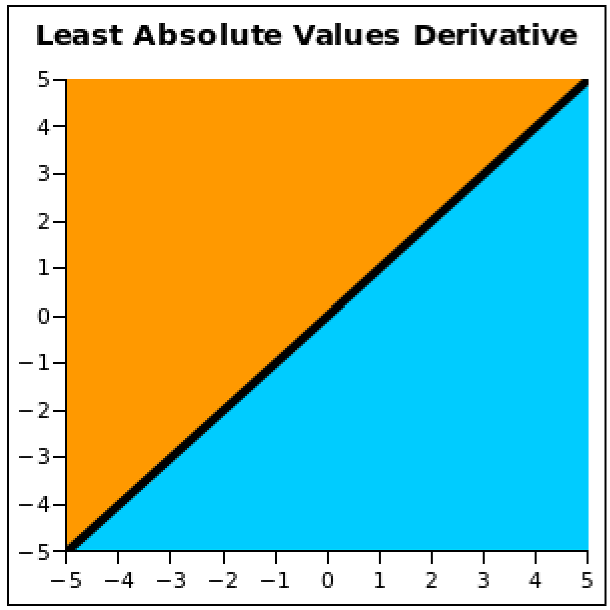
Least Absolute Values
Least Squares
LINEX
Exponential
Arctangent
Quartic
Cosine Squared
Hyperbolic Sine
Hyperbolic Tangent
Arc Hyperbolic
Normal
Funnel
Hourglass

Functions for positive y and z

Symmetric Ratio
Ratio-1
Ratio-2
Ratio-3

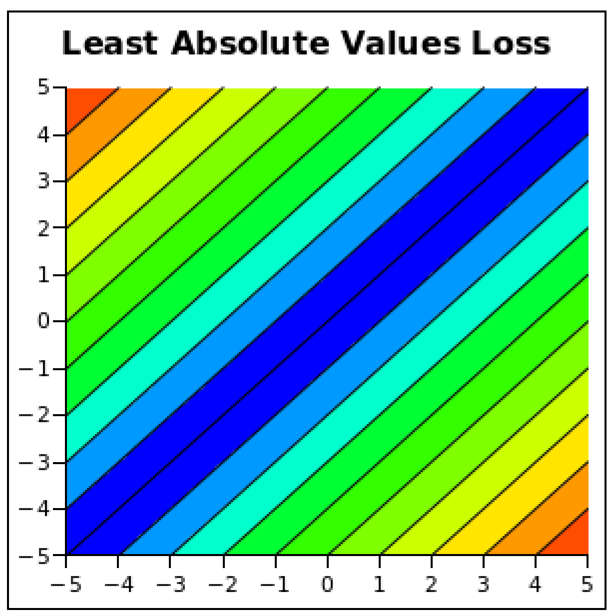
Least Absolute Values

$$\frac{\partial^2 g}{\partial z^2} = 0, \quad z \neq y$$



$$\frac{\partial g}{\partial z} = \begin{Bmatrix} -1, & z < y \\ +1, & z > y \end{Bmatrix}$$

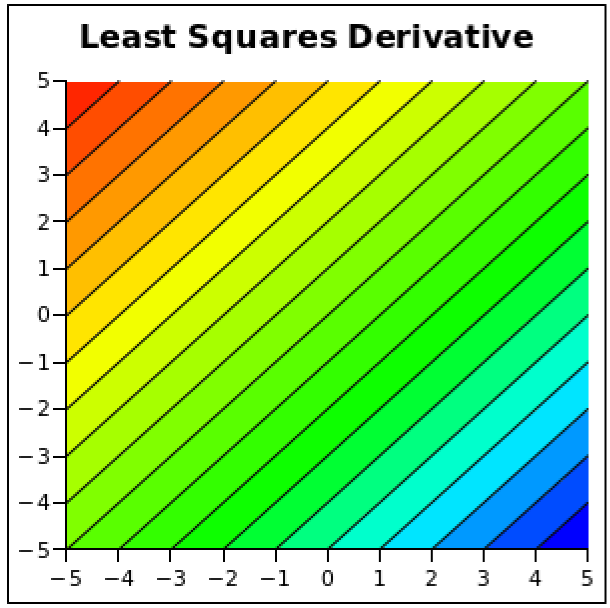
$$g = |z - y|$$

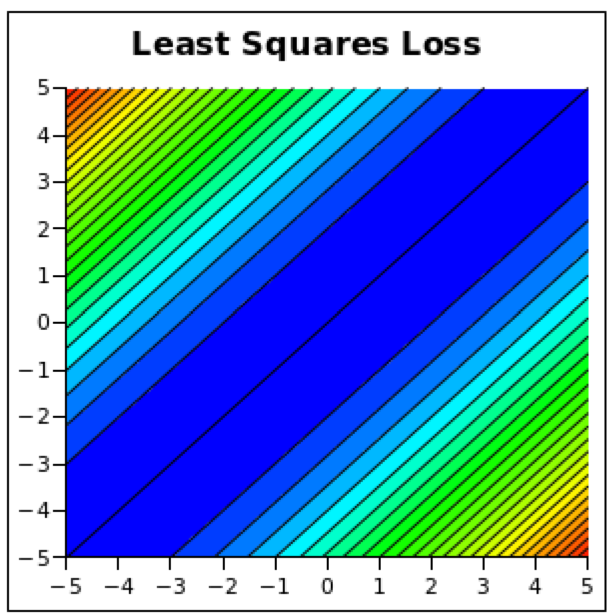


Least Squares

$$\frac{\partial^2 g}{\partial z^2} = 1 > 0$$

$$\frac{\partial g}{\partial z} = z - y$$



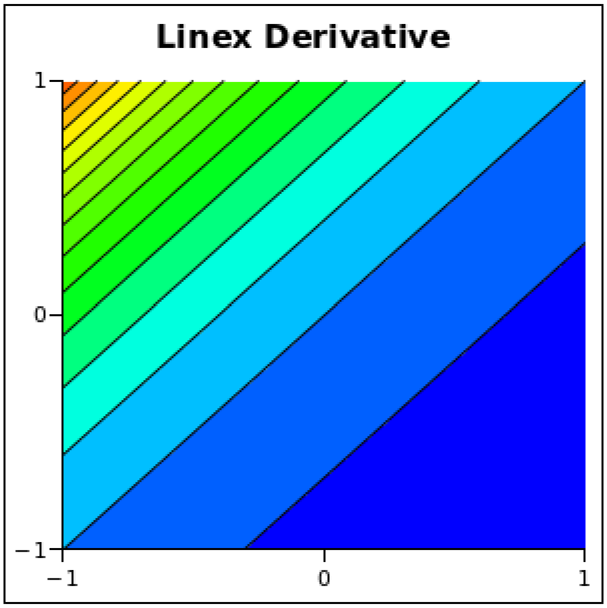


$$g = \frac{1}{2}(z - y)^2$$

LINEX

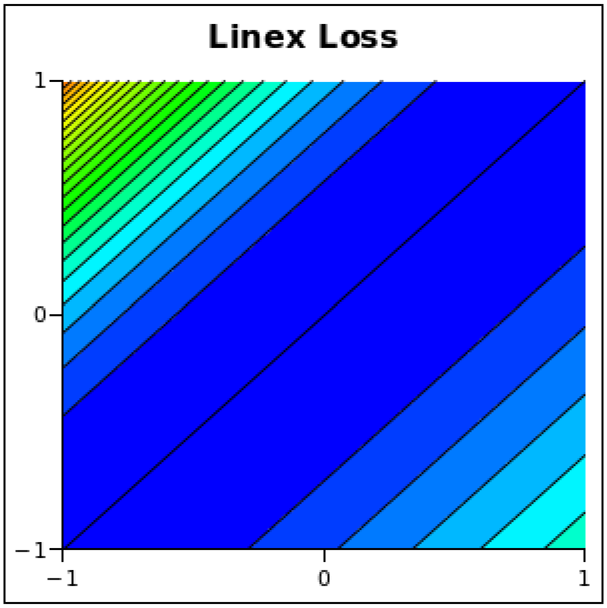
(Varian, 1975)

$$\frac{\partial^2 g}{\partial z^2} = \alpha^2 e^{\alpha(z-y)} > 0$$



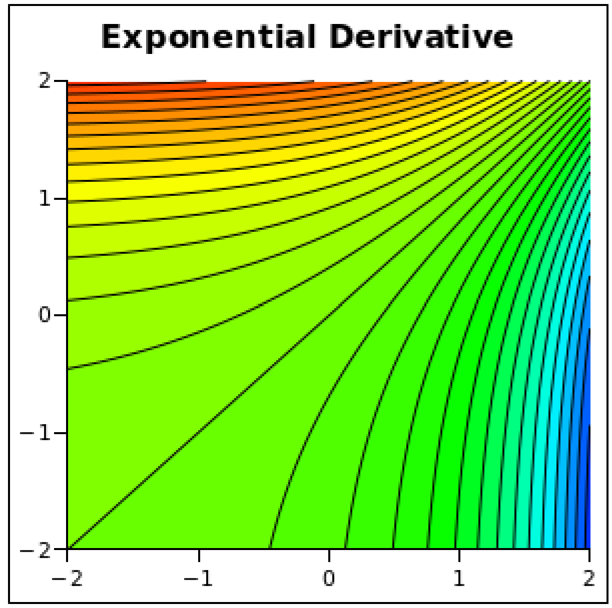
$$\frac{\partial g}{\partial z} = \alpha e^{\alpha(z-y)} - \alpha$$

$$g = e^{\alpha(z-y)} - \alpha(z-y) - 1$$



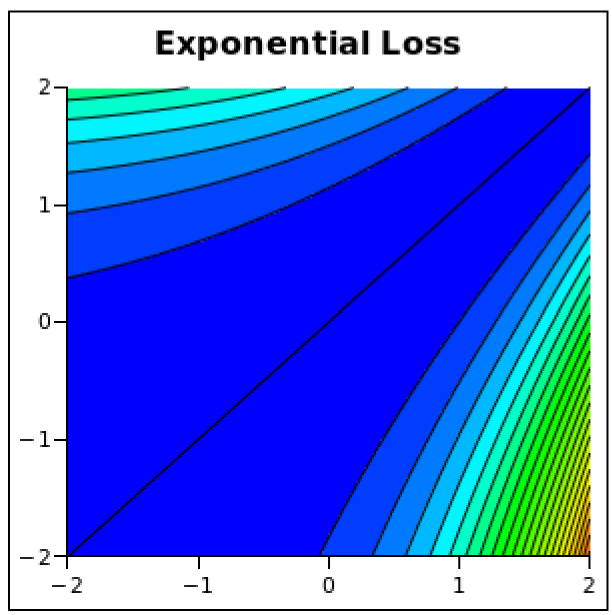
Exponential

$$\frac{\partial^2 g}{\partial z^2} = \frac{1}{p^2} e^{\frac{z}{p}} > 0$$



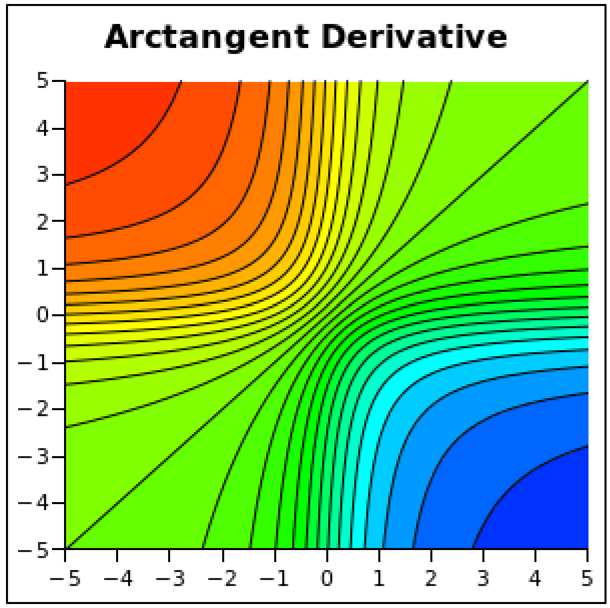
$$\frac{\partial g}{\partial z} = \frac{1}{p} \left(e^{\frac{z}{p}} - e^{\frac{y}{p}} \right)$$

$$g = e^{\frac{z}{p}} - \left(\left(\frac{z - y}{p} \right) + 1 \right) e^{\frac{y}{p}}$$



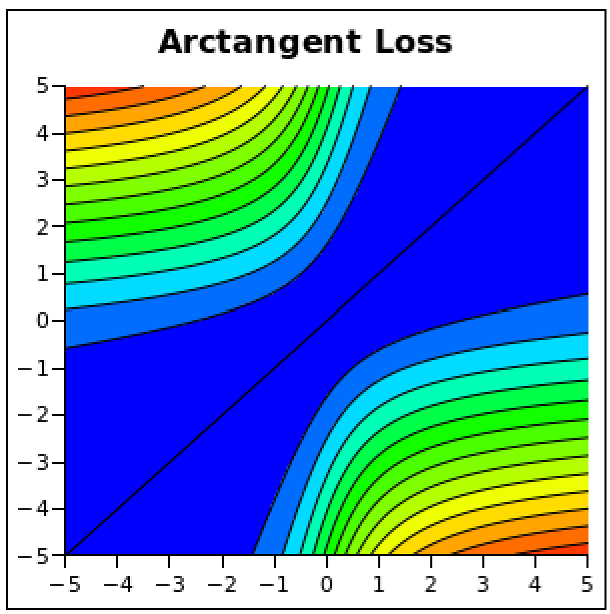
Arctangent

$$\frac{\partial^2 g}{\partial z^2} = \frac{1}{z^2+p^2} > 0$$



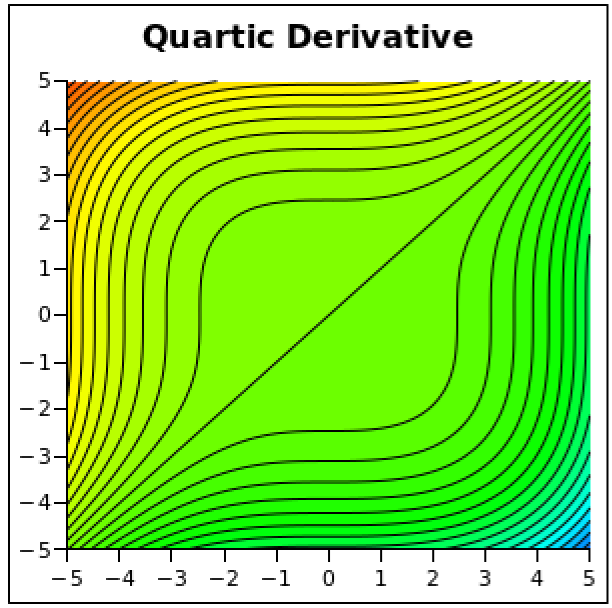
$$\frac{\partial g}{\partial z} = \frac{1}{p}\left(\arctan\left(\frac{z}{p}\right) - \arctan\left(\frac{y}{p}\right)\right)$$

$$g = \frac{z}{p}\left(\arctan\left(\frac{z}{p}\right) - \arctan\left(\frac{y}{p}\right)\right) - \frac{1}{2}\ln\left(\frac{1+\left(\frac{z}{p}\right)^2}{1+\left(\frac{y}{p}\right)^2}\right)$$



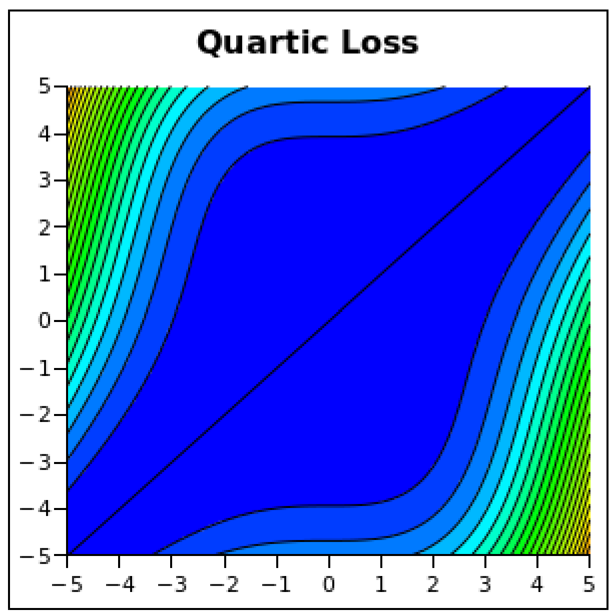
Quartic

$$\frac{\partial^2 g}{\partial z^2} = z^2 > 0$$



$$\frac{\partial g}{\partial z} = \frac{1}{3}\left(z^3 - y^3\right)$$

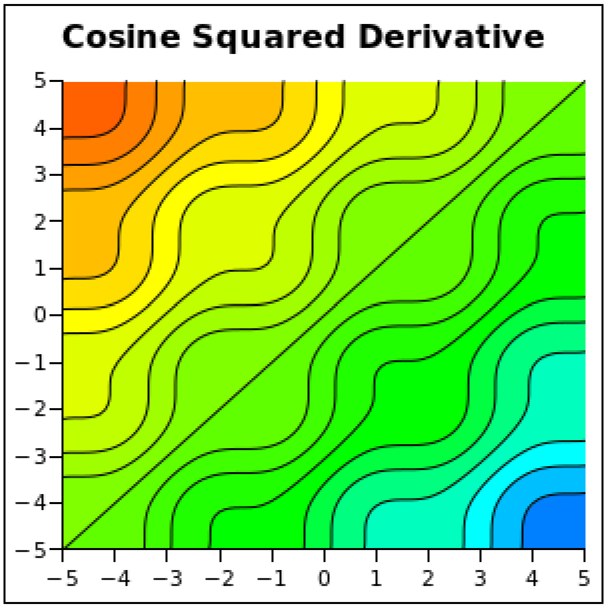
$$g = \frac{1}{12}z^4 - \frac{1}{3}y^3 z + \frac{1}{4}y^4$$



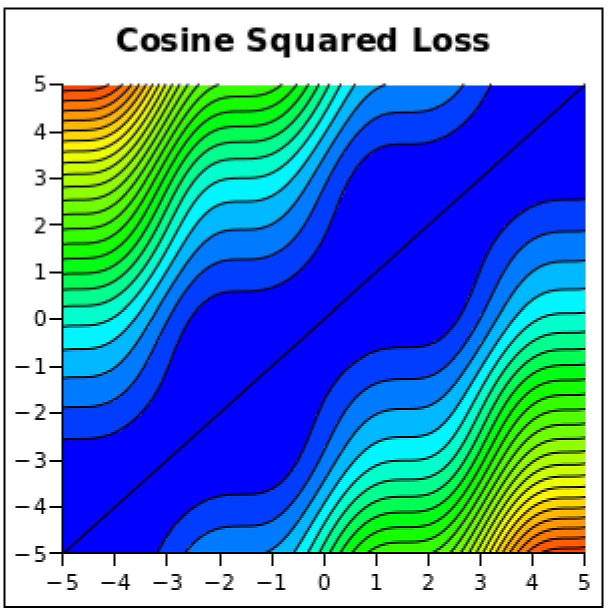
Cosine Squared

$$\frac{\partial^2 g}{\partial z^2} = \frac{1}{p^2}\left(\cos\left(\frac{z}{p}\right)\right)^2 = \frac{1}{2p}\left(1 + \cos\left(\frac{2z}{p}\right)\right)$$

$$\frac{\partial g}{\partial z} = \frac{1}{2p}\left(\left(\frac{z-y}{p}\right) + \frac{1}{2}\left(\sin\left(\frac{2z}{p}\right) - \sin\left(\frac{2y}{p}\right)\right)\right)$$

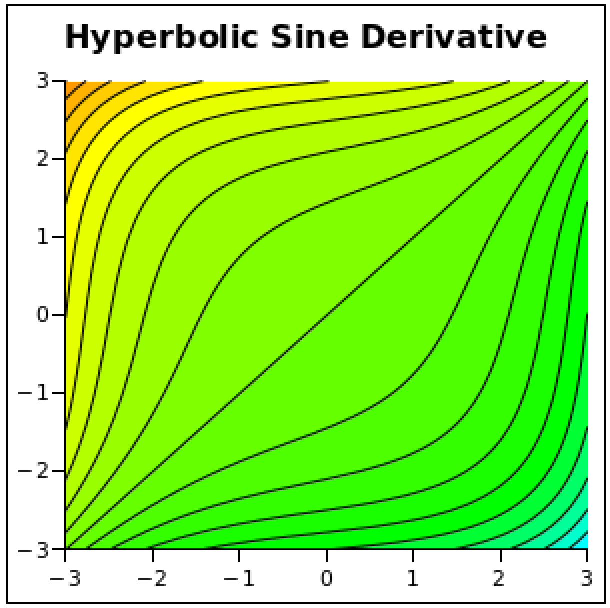


$$g = \frac{1}{4}\left(\left(\frac{z-y}{p}\right)^2 - \left(\frac{z-y}{p}\right)\sin\left(\frac{2y}{p}\right) - \frac{1}{2}\left(\cos\left(\frac{2z}{p}\right) - \cos\left(\frac{2y}{p}\right)\right)\right)$$



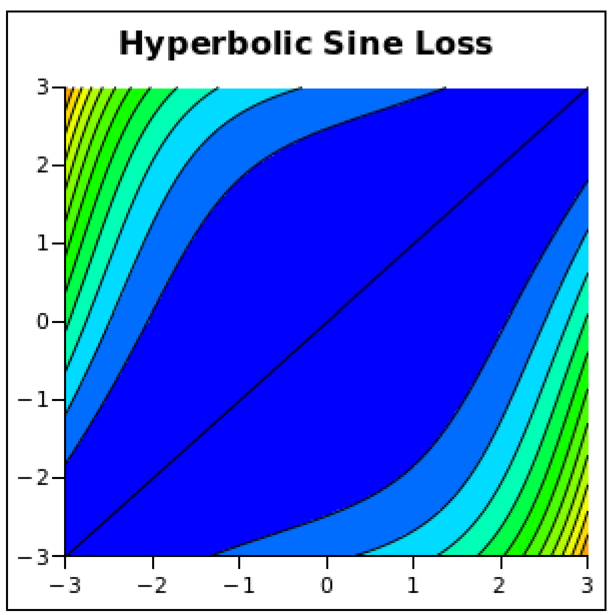
Hyperbolic Sine

$$\frac{\partial^2 g}{\partial z^2} = \frac{1}{p^2}\cosh\left(\frac{z}{p}\right) > 0$$



$$\frac{\partial g}{\partial z} = \frac{1}{p}\left(\sinh\left(\frac{z}{p}\right) - \sinh\left(\frac{y}{p}\right)\right)$$

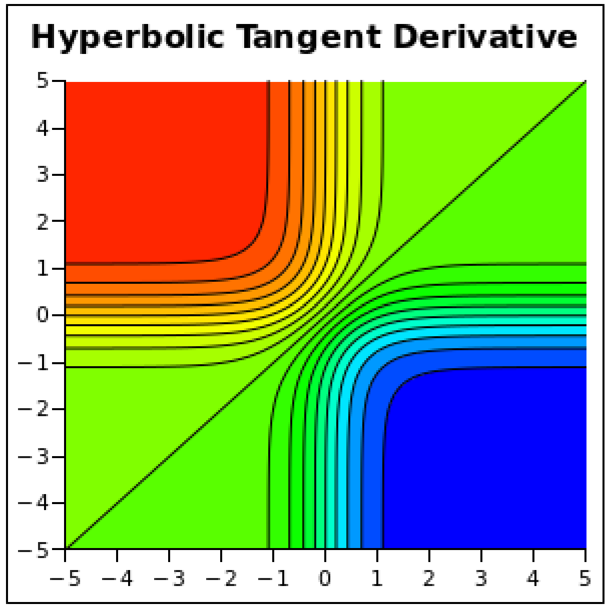
$$g = \cosh\left(\frac{z}{p}\right) - \cosh\left(\frac{y}{p}\right) - \left(\frac{z-y}{p}\right)\sinh\left(\frac{y}{p}\right)$$



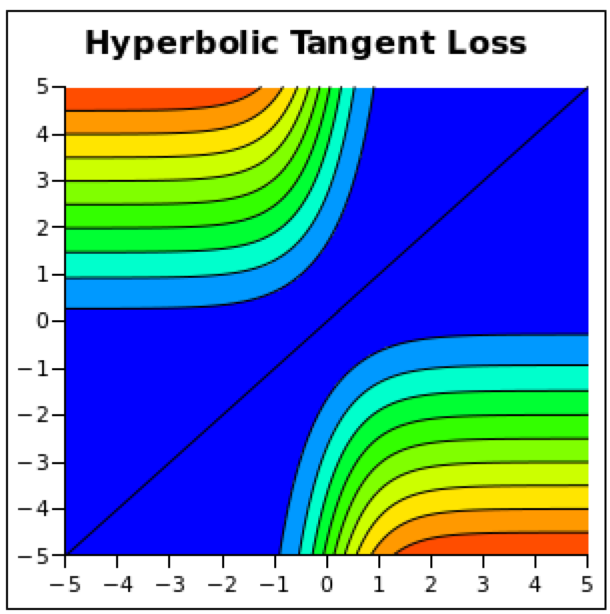
Hyperbolic Tangent

$$\frac{\partial^2 g}{\partial z^2} = \frac{1}{p^2}\left(\operatorname{sech}\left(\frac{z}{p}\right)\right)^2 > 0$$

$$\frac{\partial g}{\partial z} = \frac{1}{p}\left(\tanh\left(\frac{z}{p}\right) - \tanh\left(\frac{y}{p}\right)\right)$$

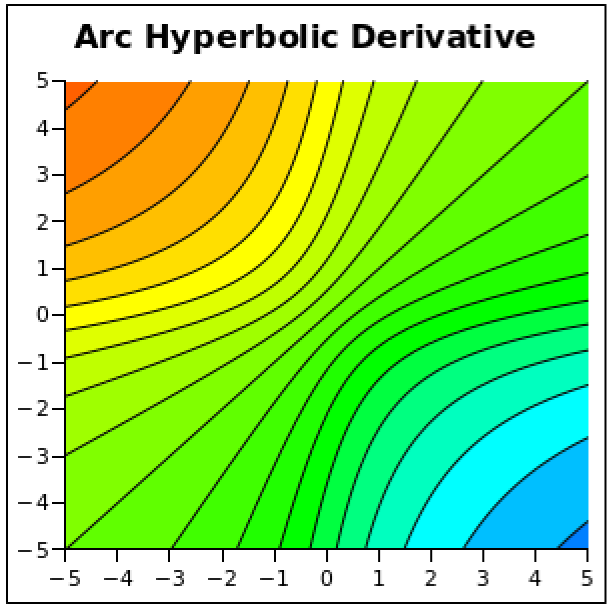


$$g = \ln\left(\frac{\cosh\left(\frac{z}{p}\right)}{\cosh\left(\frac{y}{p}\right)}\right) - \left(\frac{z-y}{p}\right)\tanh\left(\frac{y}{p}\right)$$



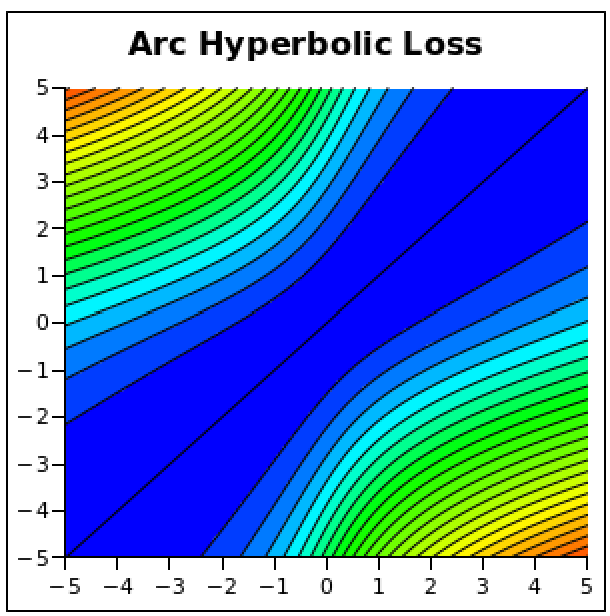
Arc Hyperbolic

$$\frac{\partial^2 g}{\partial z^2} = \frac{1}{p^2} \frac{1}{\sqrt{\left(\frac{z}{p}\right)^2 + 1}} > 0$$



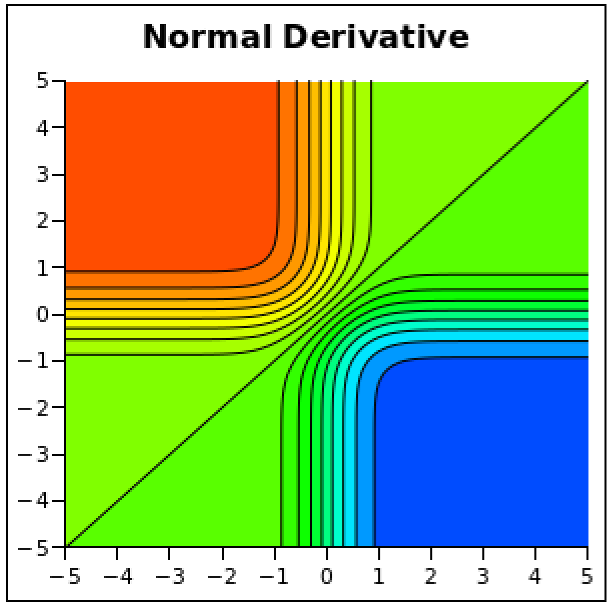
$$\frac{\partial g}{\partial z} = \frac{1}{p}\left(\operatorname{asinh}\left(\frac{z}{p}\right) - \operatorname{asinh}\left(\frac{y}{p}\right)\right)$$

$$g = \frac{z}{p}\left(\operatorname{asinh}\left(\frac{z}{p}\right) - \operatorname{asinh}\left(\frac{y}{p}\right)\right) - \left(\sqrt{\left(\frac{z}{p}\right)^2 + 1} - \sqrt{\left(\frac{y}{p}\right)^2 + 1}\right)$$



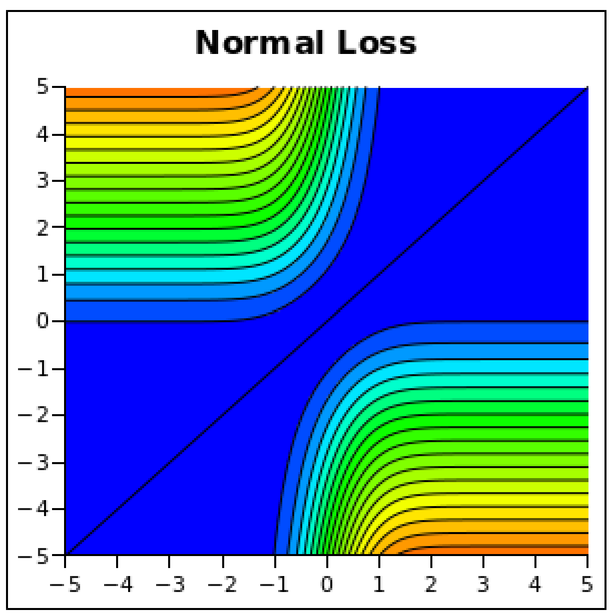
Normal

$$\frac{\partial^2 g}{\partial z^2} = \frac{1}{p^2} e^{-\left(\frac{z}{p}\right)^2} > 0$$



$$\frac{dg}{dz} = \frac{\sqrt{\pi}}{2\,p}\left(\operatorname{erf}\left(\frac{z}{p}\right) - \operatorname{erf}\left(\frac{y}{p}\right)\right)$$

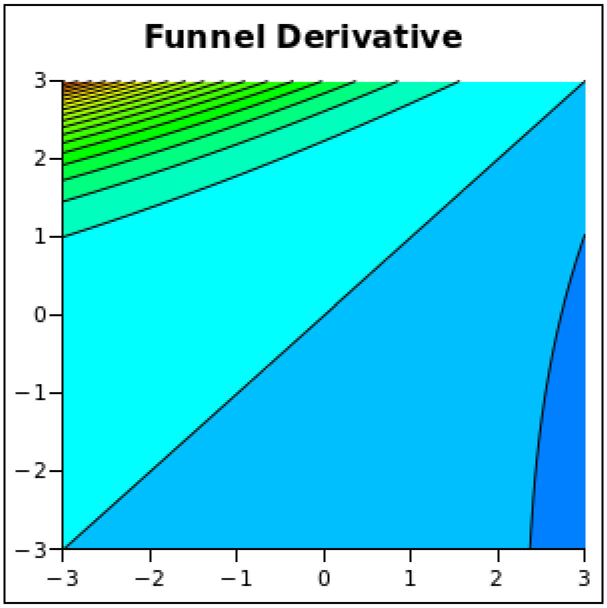
$$g = \frac{1}{2}\left(\sqrt{\pi}\,\frac{z}{p}\left(\operatorname{erf}\left(\frac{z}{p}\right) - \operatorname{erf}\left(\frac{y}{p}\right)\right) + e^{-\left(\frac{z}{p}\right)^2} - e^{-\left(\frac{y}{p}\right)^2}\right)$$



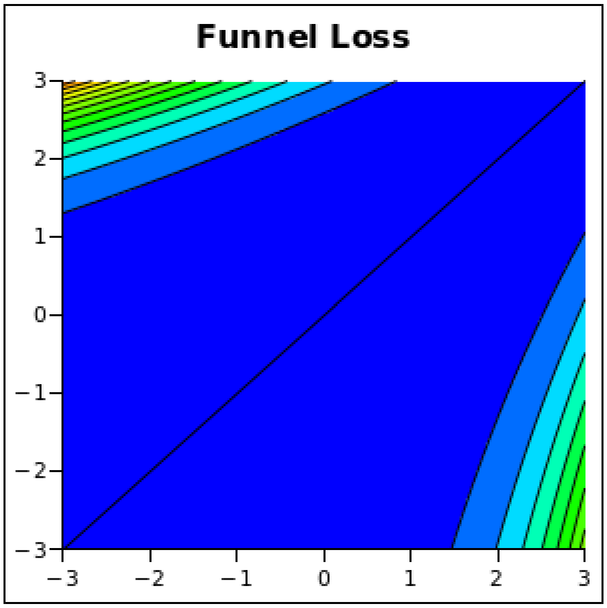
Funnel

$$\frac{\partial^2 g}{\partial z^2} = \frac{b^2}{a^4}\left(\left(\frac{z-y}{b}\right)^2 + 1\right) e^{\frac{z}{a}}$$

$$\frac{\partial g}{\partial z} = \frac{1}{a}\left[\left(\frac{z-y}{a}\right)^2 - 2\left(\frac{z-y}{a}\right) + \frac{b^2}{a^2} + 2\right] e^{\frac{z}{a}} - \frac{1}{a}\left[\frac{b^2}{a^2} + 2\right] e^{\frac{y}{a}}$$



$$g = \left[\left(\frac{z-y}{a}\right)^2 - 4\left(\frac{z-y}{a}\right) + \frac{b^2}{a^2} + 6\right] e^{\frac{z}{a}} - \left[\left(\frac{b^2}{a^2} + 2\right)\left(\frac{z-y}{a}\right) + \frac{b^2}{a^2} + 6\right] e^{\frac{y}{a}}$$

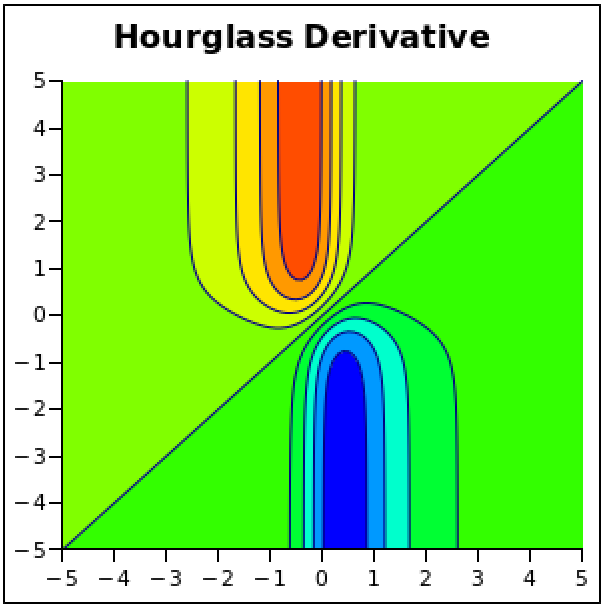


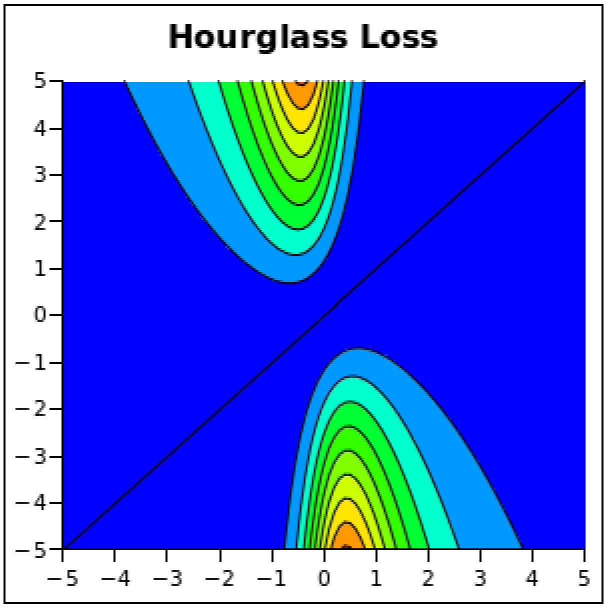
Hourglass

$$\frac{\partial^2 g}{\partial z^2} = \frac{1}{q^2} \frac{1}{\left(\frac{y}{p}\right)^2 + 1} \frac{1}{\left(\left(\frac{z}{q}\right)^2 + 1\right)^2}$$

$$\frac{\partial g}{\partial z} = \frac{1}{2q} \frac{1}{\left(\frac{y}{p}\right)^2 + 1} \left[\arctan\left(\frac{z}{q}\right) - \arctan\left(\frac{y}{q}\right) + \frac{\frac{z}{q}}{\left(\frac{z}{q}\right)^2 + 1} - \frac{\frac{y}{q}}{\left(\frac{y}{q}\right)^2 + 1}\right]$$

$$g = \frac{1}{2}\left(\frac{1}{\left(\frac{y}{p}\right)^2 + 1}\right)\left(\frac{z}{q}\left(\arctan\left(\frac{z}{q}\right) - \arctan\left(\frac{y}{q}\right)\right) - \frac{z-y}{q}\frac{\frac{y}{q}}{\left(\frac{y}{q}\right)^2 + 1}\right)$$

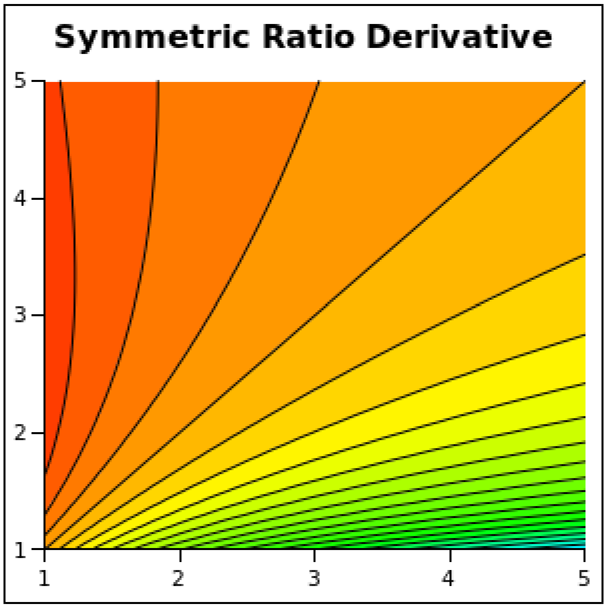




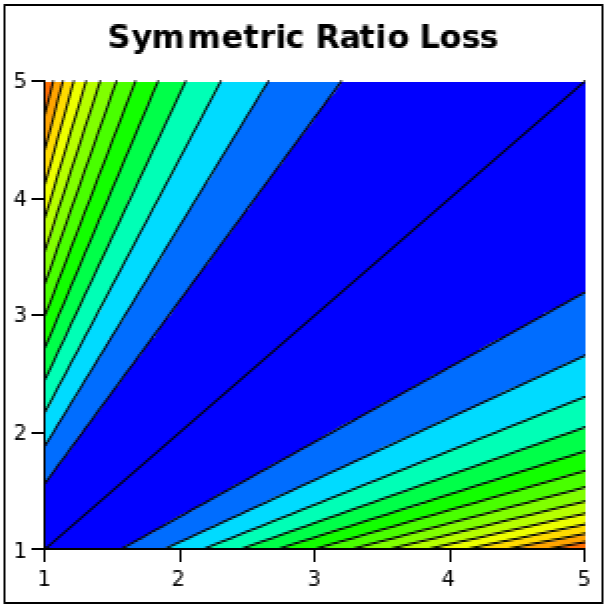
It is unnecessary to parameterize the ratio losses. In the symmetric and Ratio-2 functions, the units cancel. In Ratio-1 and Ratio-3, change of the unit of measure merely scales the result by a constant factor.

Symmetric Ratio

$$\frac{\partial^2 g}{\partial z^2} = \frac{\ln\left(\frac{y}{z}\right) + 1}{z^2} \qquad y, z > 0$$



$$\frac{\partial g}{\partial z} = \frac{1}{z} \ln\left(\frac{z}{y}\right)$$

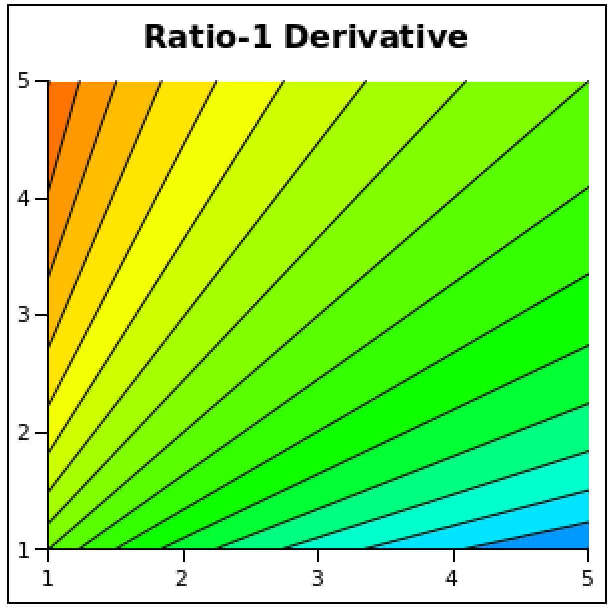


$$g = \frac{1}{2} \left(\ln\left(\frac{z}{y}\right)\right)^2$$

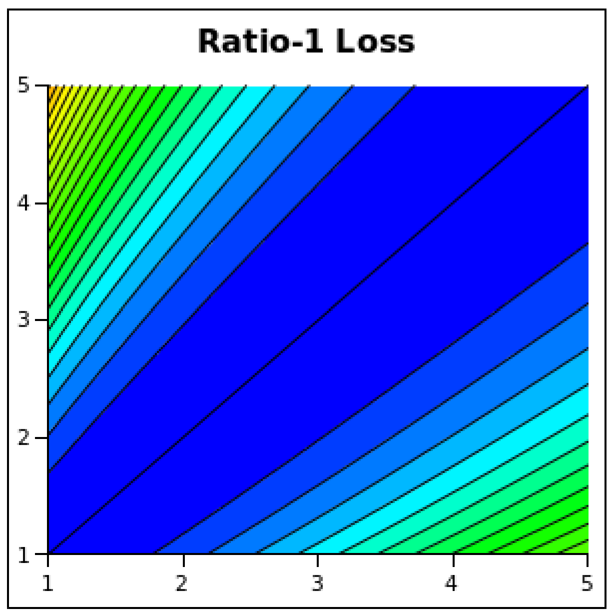
This is equivalent to the least-squares function on log-transformed variables.

Ratio-1

$$\frac{\partial^2 g}{\partial z^2} = \frac{1}{z}$$



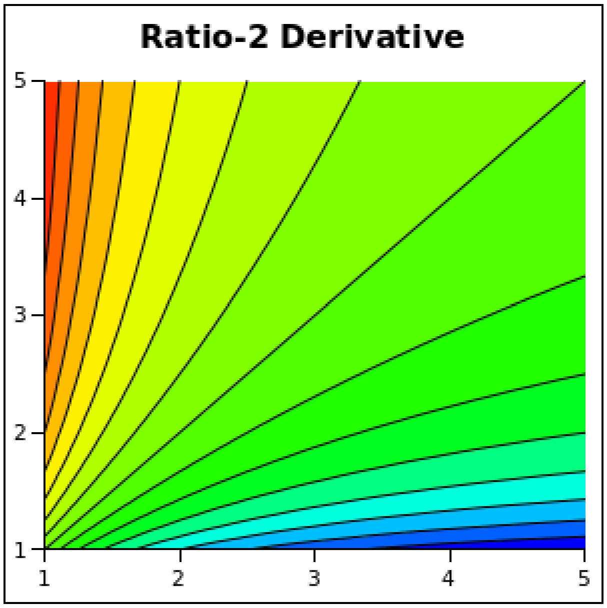
$$\frac{\partial g}{\partial z} = \ln\left(\frac{z}{y}\right)$$



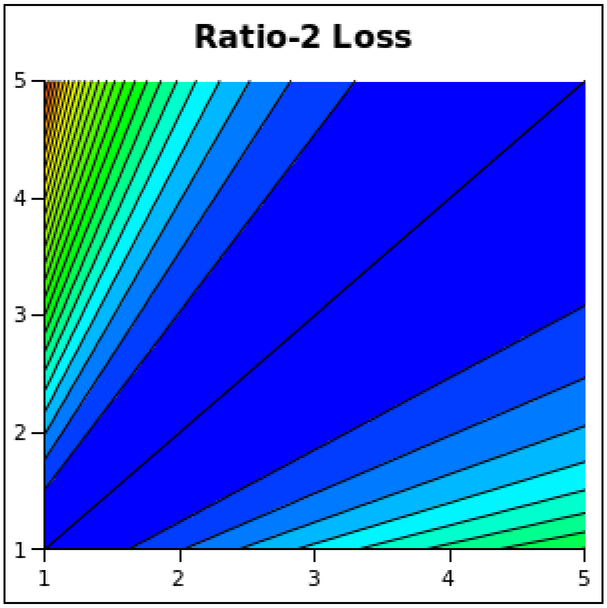
$$g = z \ln\left(\frac{z}{y}\right) - (z - y)$$

Ratio-2

$$\frac{\partial^2 g}{\partial z^2} = \frac{1}{z^2}$$



$$\frac{\partial g}{\partial z} = -\frac{1}{z} + \frac{1}{y}$$

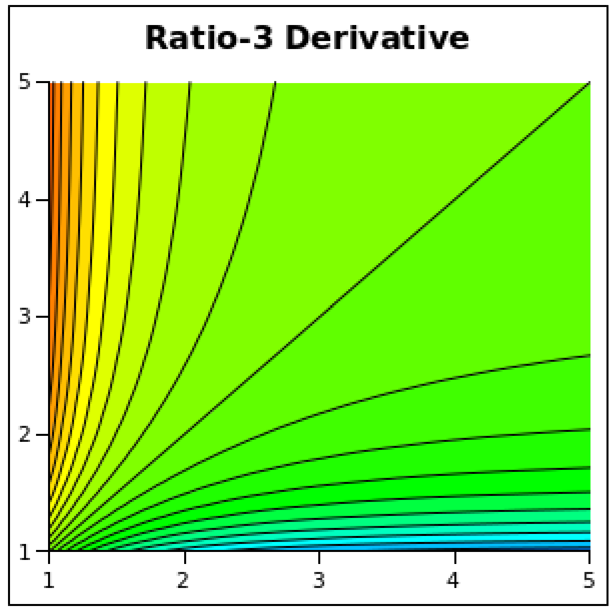


$$g = \frac{z}{y} + \ln\left(\frac{y}{z}\right) - 1$$

Interestingly, this is equivalent to the LINEX function on log-transformed variables.

Ratio-3

$$\frac{\partial^2 g}{\partial z} = z^{-3}$$



$$\frac{\partial g}{\partial z} = -\frac{1}{2} z^{-2} + \frac{1}{2} y^{-2}$$

$$g = \frac{1}{2}\frac{1}{z} + \frac{1}{2}\frac{z}{y^2} - \frac{1}{y}$$