A Bi-Symmetric Log transformation for wide-range data.

J. Beau W. Webber[1,2,3]**.**

[1] School of Physical Sciences, University of Kent, Canterbury, CT2 7NH, UK
[2] Institute of Petroleum Engineering, Heriot Watt, Edinburgh, EH14 4AS, UK
[3] Lab-Tools Ltd., Unit 6F, Thomas Way, LakesView International Business Park, Hersden, Canterbury, Kent. CT3 4JZ, UK

E-mail: `J.B.W.Webber@kent.ac.uk`

Abstract. The logarithmic transformation has long been used to present data that has both large and small components that are significant, such as neutron scattering data, or to present data that say covers a wide range of time-scales, such as NMR relaxation data. A more general transformation, that is applicable to many different disciplines, is offered here, that is particularly suitable for representing wide-range data that has both positive and negative (or zero) components. The proposed transform smoothly modifies the gradient of the transformation so that in the region near zero it remains finite. A single constant is provided to tune this behavior, so as to adjust the meaning of "region near zero". This modified logarithmic transformation can be both one-sided or symmetric, and thus can transform negative data to scaled negative data. It can be applied to both the X and Y data, when it becomes a bi-symmetric log transform.

PACS numbers: 07.05.Rm 07.05.Kf 06.30.-k 06.60.-c 06.90.+v

Keywords: data presentation, data visualization, data analysis, data management, algorithms, laboratory procedures.

Submitted to: *Meas. Sci. Technol.*

1. The Bi-Symmetric Log transformation

The logarithmic transformation has long been used to present data that has both large and small components that are significant, such as neutron scattering data, or to present data that say covers a wide range of time-scales, such as NMR relaxation data.

There have been alternative representations proposed; there is a reference on the web to "symlog" in the Python matplotlib (*Python matplotlib* 2010) (from the examples, this appears to switch abruptly between linear and logarithmic representations), and there is also a reference on the Agilent web-site to the GeneSpring Manual, listing "Symmetric Log" as a plotting option (*GeneSpring Manual* 2007) (this does not appear to accept an adjustable constant).

A more general transformation, that is applicable to many different disciplines, is offered here, that is particularly suitable for representing wide-range data that has both positive and negative (or zero) components.

The difficulty with the conventional logarithmic transformation when used as a display transformation is that near zero the transfer function tends to infinity; thus it can over-emphasize this near-zero data. Indeed it can not represent any negative or zero component in the data. One device to enable it's use is to add a constant to the data, so it never goes negative; however this now presents data that is smoothly oscillating around zero as asymmetric.

The proposed transform smoothly modifies the gradient of the transformation so that in the region near zero it remains finite, see figure 1. A single constant C is provided to tune this behavior, so as to adjust the meaning of "region near zero". The default value of this constant is $1/\ln(10)$; this gives a unity transfer function at zero (as per a straight line through the origin), but other values can be applied as wished, to focus into the region near zero or not.

This modified logarithmic transformation can be both one-sided or symmetric, and thus can transform negative data to scaled negative data. It can be applied to both the X and Y data, when it becomes a bi-symmetric log transform.

The slope (or transfer function) $k(x)$ is given by

$$\frac{d\ln(x)}{dx} = 1/x, \tag{1}$$

which tends to positive infinity at zero.

A simple, smoothly acting, inverse limiting function is then applied to this transfer function $k(x)$, such that $k(x)$ has a value of unity at $x = 0$:

$$k(x) = (1 + (x/C))^{-1} \tag{2}$$

The constant C adjusts the slope of the transfer function near the origin. Default value $1/\ln(10)$ then gives a unity slope transfer function near the origin :

$$k(x) = (1 + x\ln(10))^{-1} \tag{3}$$

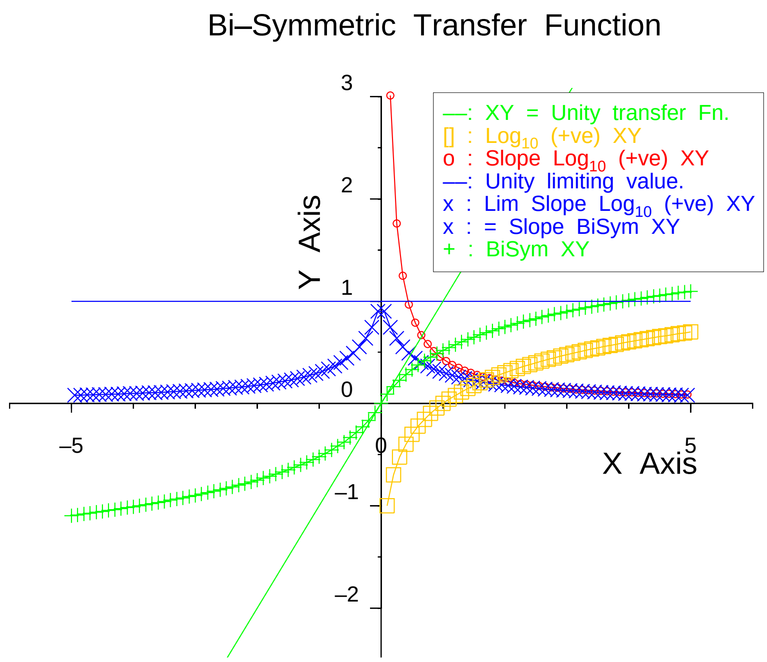

Figure 1. Demonstration of the process of converting a standard logarithm function to a Bi-Symmetric Transfer Function.
a) – A unity y=x transfer function through the origin.
b) □ A standard $y = \ln(x)$ function.
c) o The slope of the previous curve:$y = k(x) = (ln(x)) = 1/x$.
d) – A unity amplitude limiting value through $y = 1$.
e) x An inverse limited curve $k(x)$ such that
f) x $y = k(x) = (1 + x \ln(10))^{-1}$
g) + The Bi-Symmetric Transfer Function,
being the integral of the previous $k(x)$ curve, with unity slope at the origin.

This then gives the transformation :

$$y = \operatorname{sgn}(x). \log_{10}(1 + |(x/C)|) \tag{4}$$

where $\operatorname{sgn}(x)$ is the standard mathematical Sign (or Signum) function.

The inverse, power transformation, is given by :

$$x = \operatorname{sgn}(y).C.(-1 + 10^{|y|}) \tag{5}$$

These transformations are very easy to implement, particularly in array based computer languages. Supplementary material is provided expressing these transformations in the array manipulation language Apl, both as a PDF document and as an .atf programming file that can be imported into most variants of Apl.

What is less easy, with many special cases to handle, is the creation of a suitably scaled graph window. Neither Maple nor Matlab (*MATLAB, Array language, version 7.12* 2012) appear to offer such a plot representation as a default option.

Thus it is intended to make a GUI available on the web, written in the Apl (Iverson 1962) interpreter AplX (Nabavi 2001-2011), capable of producing publication quality graphs in a number of formats including postscript, pdf, jpg and png, for the user to graph their data in bi-symmetric log form.

Example Bi-symmetric-log plots are shown in figures 2 and 3.

Figure 2 shows NMR cryoporometry traces (Mitchell et al. 2008), for water/ice in a porous silica. This representation enables the NMR free induction decay (FID) of the ice around the porous silica, with a short spin-spin T_2 relaxation time of 10 μs, to clearly be viewed on the same graph as the 4 ms T_2 Carr-Purcell-Meiboom-Gill echo train for the water in the pores. The sum of the equilibrium water and ice signals on warming, extrapolated to zero time, equals the amplitude of the super-cooled water signal on cooling.

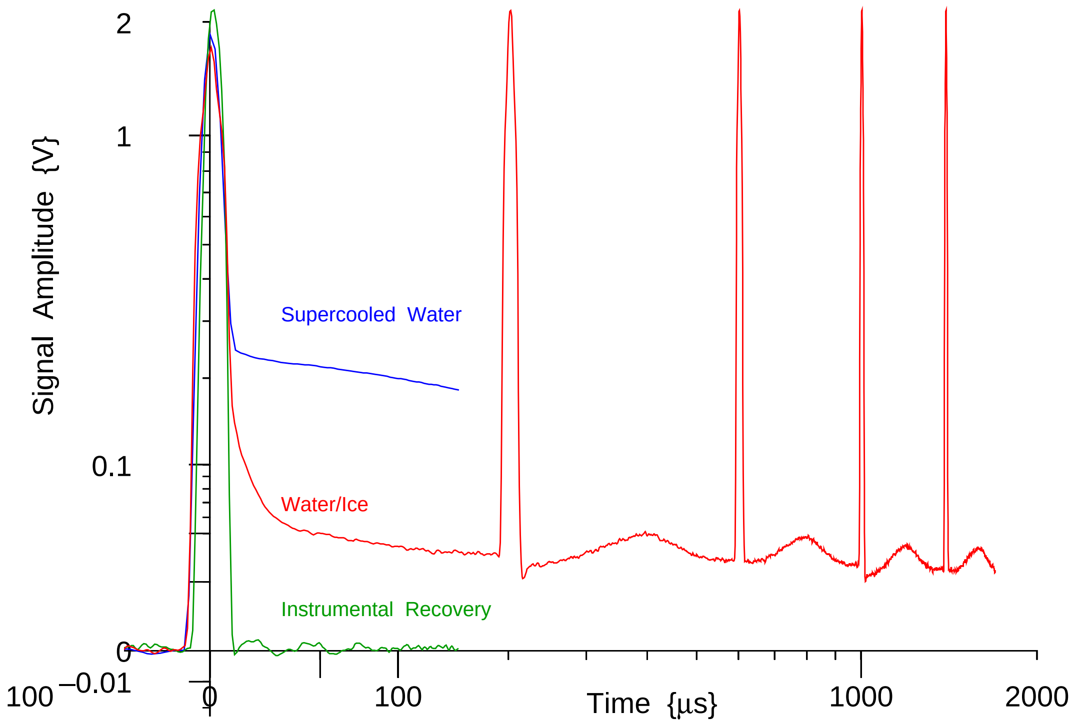

Figure 2. NMR cryoporometry signal traces for ice and water in mesoporous SBA-15 silica. Top curve: supercooled water at -5°C. Lower curve: instrumental recovery - the signal from the NMR receiver, following the powerful 90° NMR transmitter pulse, with no sample. Middle curve: signal from water/ice following freezing and re-warming to -5°C: initial 10 μs T_2 free induction decay from bulk ice around silica grains, followed by a 4 ms T_2 decay defined by peaks of echoes, from the melted water in the pores.

Figure 3 shows very wide range neutron scattering data, measured on the recently implemented neutron spectrometer NIMROD - the Near and InterMediate Range Order Diffractometer (Bowron et al. 2010) at ISIS. This figure compares the scattering from a mesoporous SBA-15 templated silica, showing intense scattering from the ordered pores in the scattering vector $q \approx 0.1\mathring{A}^{-1}$ region, together with the scattering from a clear fused silica slab, which shows more order in the $q \approx 1\mathring{A}^{-1}$ to $20\mathring{A}^{-1}$ region (Webber & Bowron in preparation). The two distinct regions have very different intensities, in very different Q ranges, yet can be clearly compared on the same graph.

2. Conclusion

This proposed graphical transformation in many cases provides a useful combination of the virtues of a linear plot and a logarithmic plot. For examples of the figure data in other representations, in figure 3: Standard log-log graph: the data at and below the x-axis cannot be displayed, but rest of the graph is fairly similar. Standard lin-lin graph: The very high dynamic range means that the graph traces just fall invisibly on the Y-axis and the X-axis.

The reliability of the data at small values of x and y depends on each case, and does require suitable values for the constant C. In figure 2 the "instrumental recovery" curve serves as guide that the other curves contain useful data for both small x and y. In figure 3, the low noise on the curves serves a similar function in y.

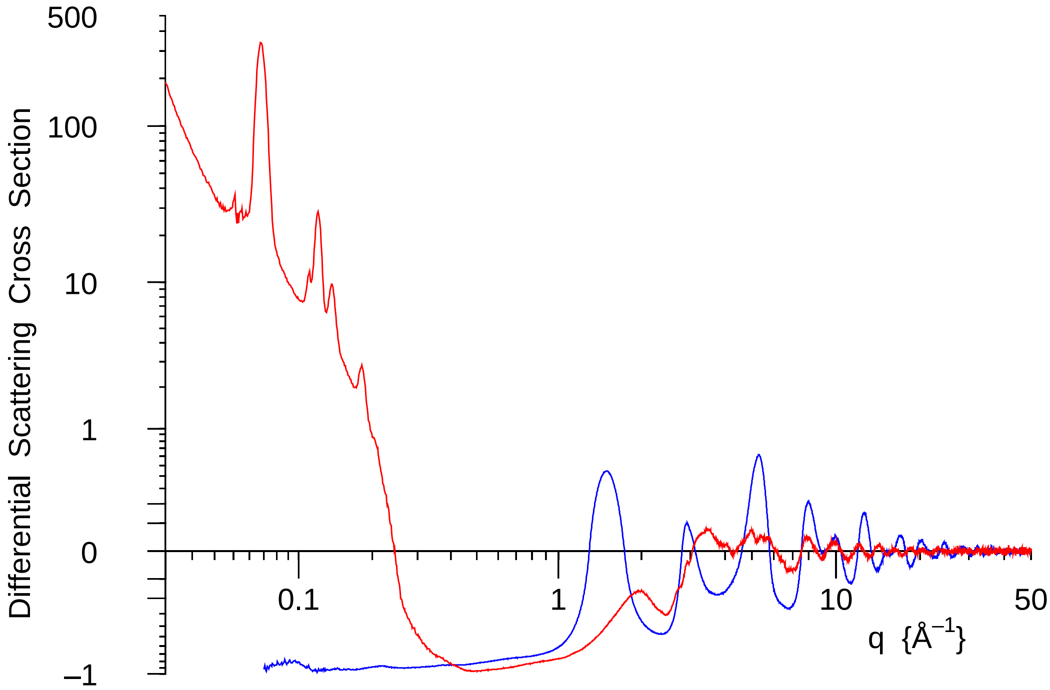

Figure 3. Differential Scattering Cross Section curves for a solid silica slab and a mesoporous SBA-15 silica.

References

Bowron D T, Soper A K, Jones K, Ansell S, Birch S, Norris J, Perrott L, Riedel D, Rhodes N J, Wakefield S R, A. Botti M A R, Grazzi F & Zoppi M 2010 *Rev. Sci. Instrum.* **81**, 033905.

GeneSpring Manual 2007.
URL: *http://www.chem.agilent.com/cag/bsp/products/gsgx/manuals/genespring-manual.pdf*

Iverson K 1962 *A Programming Language* Wiley New York.

MATLAB, Array language, version 7.12 2012.
URL: *http://www.mathworks.co.uk/*

Mitchell J, Webber J B W & Strange J 2008 *Phys. Rep.* **461**, 1–36.
URL: *http://dx.doi.org/10.1016/j.physrep.2008.02.001*

Nabavi R 2001-2011 'MicroApl's Apl Array Processing Interpreter AplX'. MicoApl's AplX interpreter.
URL: *http://www.microapl.co.uk/APL/*

Python matplotlib 2010.
URL: *http://stackoverflow.com/questions/3305865/what-is-the-difference-between-log-and-symlog*

Webber J B W & Bowron D T in preparation.