NAME

compress − Use a log transformation to compress the vertical scale of data.

SYNOPSIS

${HOME}/bin/exe/compress [K] [1=input] [2=output]

DESCRIPTION

The numerical parameter K, which defaults to 0 if it is not provided on the command line, determines the steepness of the data compression described below. If the name of the input or output file is not specified on the command line, it is prompted for and read from the keyboard. In that case the name suggested for the output file is the name of the input file suffixed by .cpr for "compressed."

The input file is read once to determine whether it contains x−y or x−y−z data, and to find *vmax*, the absolutely largest, and *vmin*, the absolutely smallest, of the vertical-axis data. Then it is rewound and read again to compress the vertical-axis data and write the output file. Comment lines in the input data (beginning # or *) are skipped. If a read error occurs the program writes a message and stops with return code 1; if the number of input data columns is not 2 or 3, the program writes a message and stops with return code 2; if the number of input columns changes, the program writes a message and stops with return code 3. If there are no data, or if their range is zero, the program writes a message and stops with return code 4.

Each non-comment record from the input file is written to the output file, with the v = y (for x−y data) or v = z (for x−y−z data) values compressed in the following way.

$$v \leftarrow sgn(v) \times \left| \frac{ln(1 + |s \times v/vmax|) - ln(1 + s \times vmin/vmax)}{ln(1 + s) - ln(1 + s \times vmin/vmax)} \right|$$

where $s = 10^K$. This scaling preserves the sign and zeros of the function, while mapping vmin to 0 and vmax to +1 (if the absolutely largest input value is positive) or −1 (if the absolutely largest input value is negative). Thus the compressed values are all in [−1,1]. Increasing the steepness s of the compression decreases the contrast between small and large values of v, which is often necessary if values near the extremes are to be distinguishable in a graph.

UNITS and FILES

0 filename prompts; error and informational messages
1 input x−y or x−y−z data
2 compressed output data
5 keyboard responses to prompts

DIAGNOSTICS

If an error occurs, an appropriate message is written and the program stops. These are the possible return codes.

0 all went well
1 an input data item could not be read
2 the input file did not contain 2 or 3 columns
3 the number of input columns changed
4 the data have zero range so cannot be compressed

AUTHOR

Michael Kupferschmid

EXAMPLE

```
unix[1] cat input
1  1.E+5
2 -1.E+3
3  1.E+1
4  0
5  1.E-5
6 -1.E-3
7  1.E-1
unix[2] compress 10 1=input 2=output
unix[3] cat output
  1.000000E+00  1.000000E+00
  2.000000E+00 -8.000000E-01
  3.000000E+00  6.000000E-01
  4.000000E+00  0.000000E+00
  5.000000E+00  3.010300E-02
  6.000000E+00 -2.004321E-01
  7.000000E+00  4.000043E-01
unix[4]
```

In this two-column input data, $vmax = 10^5$ appears on line 1 and $vmin = 0$ appears on line 4. In the output those values have been mapped to 1 and 0.