IEEE 754 Standards for Single Precision Representation

http://numericalmethods.eng.usf.edu

IEEE-754 Floating Point Standard

- Standardizes representation of floating point numbers on different computers in single and double precision.
- Standardizes representation of floating point operations on different computers.

One Great Reference

What every computer scientist (and even if you are not) should know about floating point arithmetic!


IEEE-754 Format Single Precision

32 bits for single precision

Value = \((-1)^s \times (1.m) \times 2^{(e'-127)}\)

Example#1

Represent \(-5.5834 \times 10^{10}\) as a single precision floating point number.

Example#2

Represent \(-5.5834 \times 10^{10}\) as a single precision floating point number.

http://numericalmethods.eng.usf.edu
Exponent for 32 Bit IEEE-754

8 bits would represent

\[ 0 \leq e' \leq 255 \]

Bias is 127; so subtract 127 from representation

\[-127 \leq e \leq 128\]

Exponent for Special Cases

Actual range of \( e' \)

\[ 1 \leq e' \leq 254 \]

\( e' = 0 \) and \( e' = 255 \) are reserved for special numbers

Actual range of \( e \)

\[-126 \leq e \leq 127\]

Special Exponents and Numbers

\( e' = 0 \) — all zeros

\( e' = 255 \) — all ones

<table>
<thead>
<tr>
<th>s</th>
<th>e'</th>
<th>m</th>
<th>Represents</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>all zeros</td>
<td>all zeros</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>all zeros</td>
<td>all zeros</td>
<td>-0</td>
</tr>
<tr>
<td>0</td>
<td>all ones</td>
<td>all zeros</td>
<td>∞</td>
</tr>
<tr>
<td>1</td>
<td>all ones</td>
<td>all zeros</td>
<td>-∞</td>
</tr>
<tr>
<td>0 or 1</td>
<td>all ones</td>
<td>non-zero</td>
<td>NaN</td>
</tr>
</tbody>
</table>

IEEE-754 Format

The largest number by magnitude

\[ (1.1\ldots1)_2 \times 2^{127} = 3.40 \times 10^{38} \]

The smallest number by magnitude

\[ (1.00\ldots0)_2 \times 2^{-126} = 2.18 \times 10^{-38} \]

Machine epsilon

\( \varepsilon_{mach} = 2^{-23} = 1.19 \times 10^{-7} \)

Additional Resources

For all resources on this topic such as digital audiovisual lectures, primers, textbook chapters, multiple-choice tests, worksheets in MATLAB, Mathematica, MathCad and Maple, blogs, related physical problems, please visit

http://numericalmethods.eng.usf.edu/topics/floatingpoint_representation.html

THE END

http://numericalmethods.eng.usf.edu