IEEE floating-point in Java

Peter Sestoft, IT University of Copenhagen
sestoft@itu.dk • 2009-01-05

1 Results

Results are as required by the IEEE 754-2008 standard.
Source code in java/Numbers.java.

1.1 Arithmetic operators

+	-Inf	-2.0	-0.0	0.0	2.0	+Inf	NaN
-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	NaN	NaN
-2.0	-Inf	-4.0	-2.0	-2.0	0.0	+Inf	NaN
-0.0	-Inf	-2.0	-0.0	0.0	2.0	+Inf	NaN
0.0	-Inf	-2.0	0.0	0.0	2.0	+Inf	NaN
2.0	-Inf	0.0	2.0	2.0	4.0	+Inf	NaN
+Inf	NaN	+Inf	+Inf	+Inf	+Inf	+Inf	NaN
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

-	-Inf	-2.0	-0.0	0.0	2.0	+Inf	NaN
-Inf	NaN	-Inf	-Inf	-Inf	-Inf	-Inf	NaN
-2.0	+Inf	0.0	-2.0	-2.0	-4.0	-Inf	NaN
-0.0	+Inf	2.0	0.0	-0.0	-2.0	-Inf	NaN
0.0	+Inf	2.0	0.0	0.0	-2.0	-Inf	NaN
2.0	+Inf	4.0	2.0	2.0	0.0	-Inf	NaN
+Inf	+Inf	+Inf	+Inf	+Inf	+Inf	NaN	NaN
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

*	-Inf	-2.0	-0.0	0.0	2.0	+Inf	NaN
-Inf	+Inf	+Inf	NaN	NaN	-Inf	-Inf	NaN
-2.0	+Inf	4.0	0.0	-0.0	-4.0	-Inf	NaN
-0.0	NaN	0.0	0.0	-0.0	-0.0	NaN	NaN
0.0	NaN	-0.0	-0.0	0.0	0.0	NaN	NaN
2.0	-Inf	-4.0	-0.0	0.0	4.0	+Inf	NaN
+Inf	-Inf	-Inf	NaN	NaN	+Inf	+Inf	NaN
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

/	-Inf	-2.0	-0.0	0.0	2.0	+Inf	NaN
-Inf	NaN	+Inf	+Inf	-Inf	-Inf	NaN	NaN
-2.0	0.0	1.0	+Inf	-Inf	-1.0	-0.0	NaN
-0.0	0.0	0.0	NaN	NaN	-0.0	-0.0	NaN
0.0	-0.0	-0.0	NaN	NaN	0.0	0.0	NaN
2.0	-0.0	-1.0	-Inf	+Inf	1.0	0.0	NaN
+Inf	NaN	-Inf	-Inf	+Inf	+Inf	NaN	NaN
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

%	-Inf	-2.0	-0.0	0.0	2.0	+Inf	NaN
-Inf	NaN	NaN	NaN	NaN	NaN	NaN	NaN
-2.0	-2.0	-0.0	NaN	NaN	-0.0	-2.0	NaN
-0.0	-0.0	-0.0	NaN	NaN	-0.0	-0.0	NaN
0.0	0.0	0.0	NaN	NaN	0.0	0.0	NaN
2.0	2.0	0.0	NaN	NaN	0.0	2.0	NaN
+Inf	NaN	NaN	NaN	NaN	NaN	NaN	NaN
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

1.2 Comparison operators

==	-Inf	-2.0	-0.0	0.0	2.0	+Inf	NaN
-Inf	true	false	false	false	false	false	false
-2.0	false	true	false	false	false	false	false
-0.0	false	false	true	true	false	false	false
0.0	false	false	true	true	false	false	false
2.0	false	false	false	false	true	false	false
+Inf	false	false	false	false	false	true	false
NaN	false	false	false	false	false	false	false

!=	-Inf	-2.0	-0.0	0.0	2.0	+Inf	NaN
-Inf	false	true	true	true	true	true	true
-2.0	true	false	true	true	true	true	true
-0.0	true	true	false	false	true	true	true
0.0	true	true	false	false	true	true	true
2.0	true	true	true	true	false	true	true
+Inf	true	true	true	true	true	false	true
NaN	true	true	true	true	true	true	true

<	-Inf	-2.0	-0.0	0.0	2.0	+Inf	NaN
-Inf	false	true	true	true	true	true	false
-2.0	false	false	true	true	true	true	false
-0.0	false	false	false	false	true	true	false
0.0	false	false	false	false	true	true	false
2.0	false	false	false	false	false	true	false
+Inf	false	false	false	false	false	false	false
NaN	false	false	false	false	false	false	false

<=	-Inf	-2.0	-0.0	0.0	2.0	+Inf	NaN
-Inf	true	true	true	true	true	true	false
-2.0	false	true	true	true	true	true	false
-0.0	false	false	true	true	true	true	false
0.0	false	false	true	true	true	true	false
2.0	false	false	false	false	true	true	false
+Inf	false	false	false	false	false	true	false
NaN	false	false	false	false	false	false	false

>	-Inf	-2.0	-0.0	0.0	2.0	+Inf	NaN
-Inf	false	false	false	false	false	false	false
-2.0	true	false	false	false	false	false	false
-0.0	true	true	false	false	false	false	false
0.0	true	true	false	false	false	false	false
2.0	true	true	true	true	false	false	false
+Inf	true	true	true	true	true	false	false
NaN	false	false	false	false	false	false	false

>=	-Inf	-2.0	-0.0	0.0	2.0	+Inf	NaN
-Inf	true	false	false	false	false	false	false
-2.0	true	true	false	false	false	false	false
-0.0	true	true	true	true	false	false	false
0.0	true	true	true	true	false	false	false
2.0	true	true	true	true	true	false	false
+Inf	true	true	true	true	true	true	false
NaN	false	false	false	false	false	false	false

1.3 Two-argument mathematical functions

Math.atan2	-Inf	-2.0	-0.0	0.0	2.0	+Inf	NaN
-Inf	-2.356	-1.571	-1.571	-1.571	-1.571	-0.785	NaN
-2.0	-3.142	-2.356	-1.571	-1.571	-0.785	-0.000	NaN
-0.0	-3.142	-3.142	-3.142	-0.000	-0.000	-0.000	NaN
0.0	3.142	3.142	3.142	0.000	0.000	0.000	NaN
2.0	3.142	2.356	1.571	1.571	0.785	0.000	NaN
+Inf	2.356	1.571	1.571	1.571	1.571	0.785	NaN
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

The `IEEERemainder` function agrees with floating-point remainder `x%y` on the arguments shown here, but not in general; for instance, `IEEERemainder(7,2)` is -1 whereas `7%2` is $+1$.

Math.IEEEremainder	-Inf	-2.0	-0.0	0.0	2.0	+Inf	NaN
-Inf	NaN	NaN	NaN	NaN	NaN	NaN	NaN
-2.0	-2.0	-0.0	NaN	NaN	-0.0	-2.0	NaN
-0.0	-0.0	-0.0	NaN	NaN	-0.0	-0.0	NaN
0.0	0.0	0.0	NaN	NaN	0.0	0.0	NaN
2.0	2.0	0.0	NaN	NaN	0.0	2.0	NaN
+Inf	NaN	NaN	NaN	NaN	NaN	NaN	NaN
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Math.max	-Inf	-2.0	-0.0	0.0	2.0	+Inf	NaN
-Inf	-Inf	-2.0	-0.0	0.0	2.0	+Inf	NaN
-2.0	-2.0	-2.0	-0.0	0.0	2.0	+Inf	NaN
-0.0	-0.0	-0.0	-0.0	0.0	2.0	+Inf	NaN
0.0	0.0	0.0	0.0	0.0	2.0	+Inf	NaN
2.0	2.0	2.0	2.0	2.0	2.0	+Inf	NaN
+Inf	+Inf	+Inf	+Inf	+Inf	+Inf	+Inf	NaN
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Math.min	-Inf	-2.0	-0.0	0.0	2.0	+Inf	NaN
-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	NaN
-2.0	-Inf	-2.0	-2.0	-2.0	-2.0	-2.0	NaN
-0.0	-Inf	-2.0	-0.0	-0.0	-0.0	-0.0	NaN
0.0	-Inf	-2.0	-0.0	0.0	0.0	0.0	NaN
2.0	-Inf	-2.0	-0.0	0.0	2.0	2.0	NaN
+Inf	-Inf	-2.0	-0.0	0.0	2.0	+Inf	NaN
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Math.pow	-Inf	-2.0	-0.0	0.0	2.0	+Inf	NaN
-Inf	0.0	0.0	1.0	1.0	+Inf	+Inf	NaN
-2.0	0.0	0.25	1.0	1.0	4.0	+Inf	NaN
-0.0	+Inf	+Inf	1.0	1.0	0.0	0.0	NaN
0.0	+Inf	+Inf	1.0	1.0	0.0	0.0	NaN
2.0	0.0	0.25	1.0	1.0	4.0	+Inf	NaN
+Inf	0.0	0.0	1.0	1.0	+Inf	+Inf	NaN
NaN	NaN	NaN	1.0	1.0	NaN	NaN	NaN

1.4 One-argument mathematical functions

	-Inf	-2.000	-0.000	0.000	2.000	+Inf	NaN
Math.abs	+Inf	2.000	0.000	0.000	2.000	+Inf	NaN
Math.acos	NaN	NaN	1.571	1.571	NaN	NaN	NaN
Math.asin	NaN	NaN	-0.000	0.000	NaN	NaN	NaN
Math.atan	-1.571	-1.107	-0.000	0.000	1.107	1.571	NaN
Math.ceil	-Inf	-2.000	-0.000	0.000	2.000	+Inf	NaN
Math.cbrt	-Inf	-1.260	-0.000	0.000	1.260	+Inf	NaN
Math.cos	NaN	-0.416	1.000	1.000	-0.416	NaN	NaN
Math.exp	0.000	0.135	1.000	1.000	7.389	+Inf	NaN
Math.floor	-Inf	-2.000	-0.000	0.000	2.000	+Inf	NaN
Math.log	NaN	NaN	-Inf	-Inf	0.693	+Inf	NaN
Math.log10	NaN	NaN	-Inf	-Inf	0.301	+Inf	NaN
Math.rint	-Inf	-2.000	-0.000	0.000	2.000	+Inf	NaN
Math.sin	NaN	-0.909	-0.000	0.000	0.909	NaN	NaN
Math.signum	-1.000	-1.000	-0.000	0.000	1.000	1.000	NaN
Math.sqrt	NaN	NaN	-0.000	0.000	1.414	+Inf	NaN
Math.tan	NaN	2.185	-0.000	0.000	-2.185	NaN	NaN