3. Round-off Error and Computer Arithmetic

MAPLE. See computernumbers.mw and/or computernumbers.pdf

There are problems in using a finite subset of the reals (rational numbers, actually) to represent all reals.

Suppose each machine number can be represented as a k-digit decimal machine number in floating-point form as

$$\pm 0.d_1 d_2 \cdots d_k \times 10^n, \quad 1 \le d_1 \le 9, \quad 0 \le d_i \le 9, \quad i = 2, \ldots, k.$$

Suppose $y = 0.d_1 d_2 \cdots d_k d_{k+1} d_{k+2} \times 10^n$ is in the numerical range of the machine.

Two methods for getting the floating-point form of k digits $fl(y)$:

(1) $fl(y) = 0.d_1 d_2 \cdots d_k \times 10^n$ is chopping.

(2) rounding: add $5 \times 10^{n-(k+1)}$ and then chop to get $0.\delta_1 \delta_2 \cdots \delta_k \times 10^m$.

We refer to both cases as round-off error.

EXAMPLE (of rounding). Suppose we wish to round $.372648 \times 10^3$ and $.372653 \times 10^3$ with $k = 4$ and $n = 3$, so

$$5 \times 10^{n-(k+1)} = 5 \times 10^{-2} = 5 \times 10^{-5} \times 10^3 = .00005 \times 10^3.$$

$.372648 \times 10^3$	$.372653 \times 10^3$
$\underline{.000050 \times 10^3}$	$\underline{.000050 \times 10^3}$
$.372698 \times 10^3$	$.372703 \times 10^3$
Now chop:	
$.3726 \times 10^3$	$.3727 \times 10^3$ $\square$