Toward a Better Sine Wave
(Improving Numerical Approximation Formulae using Genetic Programming)

Matthew Streeter
Lee A. Becker
Worcester Polytechnic Institute

Motivations

- Approximations have value in formal mathematics and industrial settings
- Discovery of approximation formulae requires human insight or numerical analysis technique (f.e. Taylor series, Padé approximations)
- Genetic programming provides *general* automated technique with potential improvement over existing methods

Evaluating Approximations

- Given cost and error, utility function assigns value to an approximation
- *Reasonable* utility function assigns higher value to approximations with lower error and cost
- Pareto front represents set of approximations which are best under some reasonable utility function

Summary of Previous Work

- Discovered variations on first three terms of asymptotic expansion for Harmonic number series
- Evolved rational polynomial approximations to functions of a single variable which are superior to corresponding Padé approximations
- Evolved rational polynomial approximations for functions of multiple variables to which Padé approximation technique cannot be applied
- Refined approximations through evolution of approximations to their error function

Building upon Existing Approximations

- Seed initial population with LISP expressions corresponding to known approximations
- Seed individuals treated as available root elements in random tree generation algorithm
- One available seed individual comprises ~14.3% of initial population

Experimental Approach

- Calculated cost of each expression as sum of fixed cost associated with each primitive operator
- GP system returns Pareto front with respect to cost and error for entire population histories of one or more runs
- Used parameter settings from (Koza 1992) including populations size of 500, but with tournament selection (tournament size = 3) and generation limit of 100
- Each experiment was composed of 50 independent runs

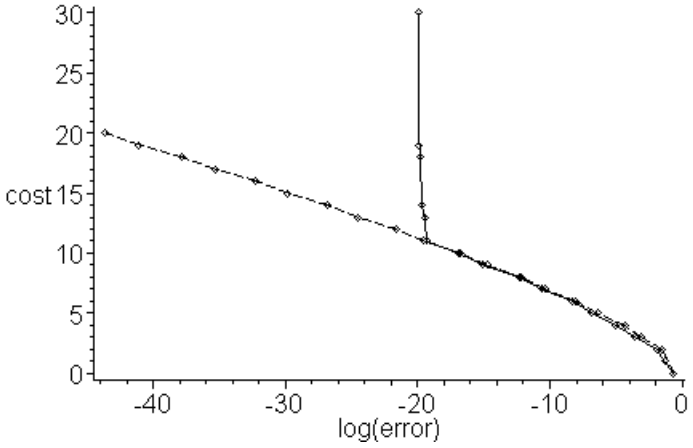
Refining Sine Wave Approximations

- Calculated Pareto front for all Padé approximations with numerator and denominator degree <= 20; this yielded 20 unique expressions
- Ran separate experiment using each of 20 Padé approximations as seed individual
- All results re-evaluated through Maple cost and error procedures
- Final result is Pareto front for all 20 experiments with respect to Maple cost and error values

Results

- Evolved refinements are typically ~twice as accurate as original approximations, given same level of cost
- Refinements represent substantial changes, not just tweaking of constants
- Accuracy of evolved approximations limited by use of long double (80-bit) floating-point arithmetic

COST	PADÉ ERROR (E_p)	EVOLVED ERROR (E_e)	E_p/E_e
10	.17196709e-16	.11906981e-16	1.4442543
9	.17246431e-14	.89135991e-15	1.9348448
8	.6488747e-12	.52074457e-12	1.2460491
7	.4269605e-10	.22289908e-10	1.9154879
6	.97933708e-8	.56504028e-8	1.7332164
5	.37808176e-6	.15708951e-6	2.4067919
4	.44176819e-4	.95235444e-5	4.6386951
3	.83028076e-3	.29332169e-3	2.8306149
2	.032510528	.012285190	2.6463187
1	NA	.047321306	NA
0	.22923809	.22923809	1.0



Future Work

- Larger population size
- Multi-objective fitness measure
- More recent GP features: automatically defined functions, architecture-altering operations
- Refinment of approximations other than Padé, e.g. Remez approximations