CREST Open Workshop 3-4th December 2018

Genetic Improvement by Evolving Program Data

W. B. Langdon
Computer Science, University College London

GI 2019

GI 2019, Montreal, ICSE-2019 workshop

Simple blue example of grammar based Genetic Improvement
opencv_gp.tar.gz
RN/18/06

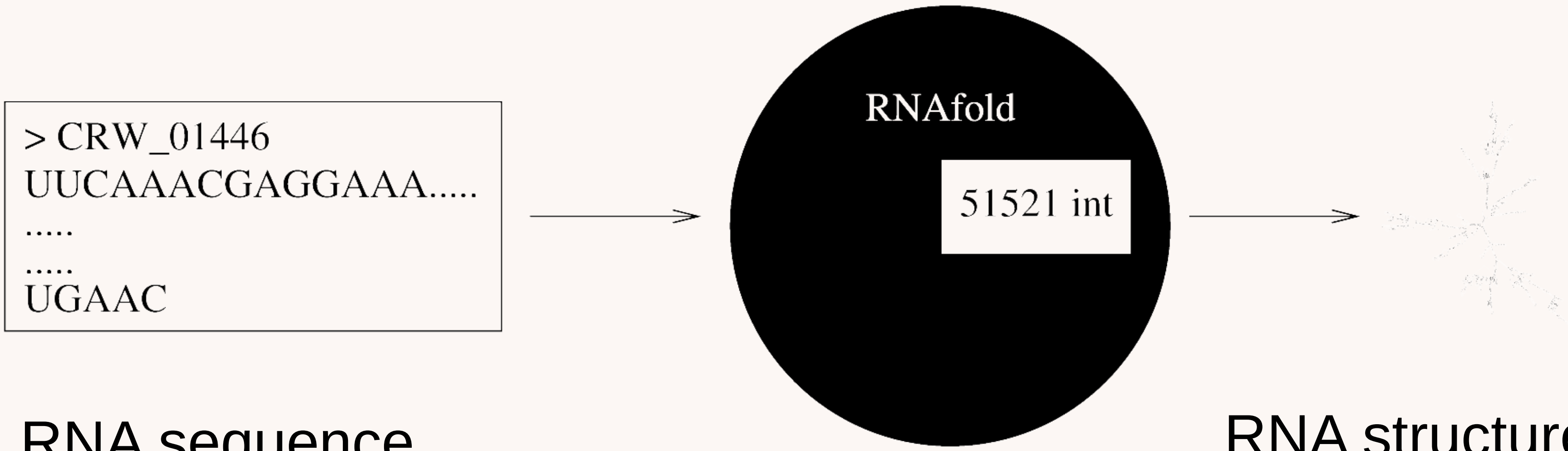

WIKIPEDIA
Genetic Improvement

CREST

Maintaining Embedded Constants

- EuroGP 2018
  - RNAfold 7000 lines of code 50000 numbers
  - On average better predictions of RNA folding.
  - Shipped since 2.4.7
- SSBSE-2018 sqrt converted to cube root
  - New functionality, double precision accuracy
- RN/18/05 generate $\log_2$ from existing open source maths framework

RNAfold

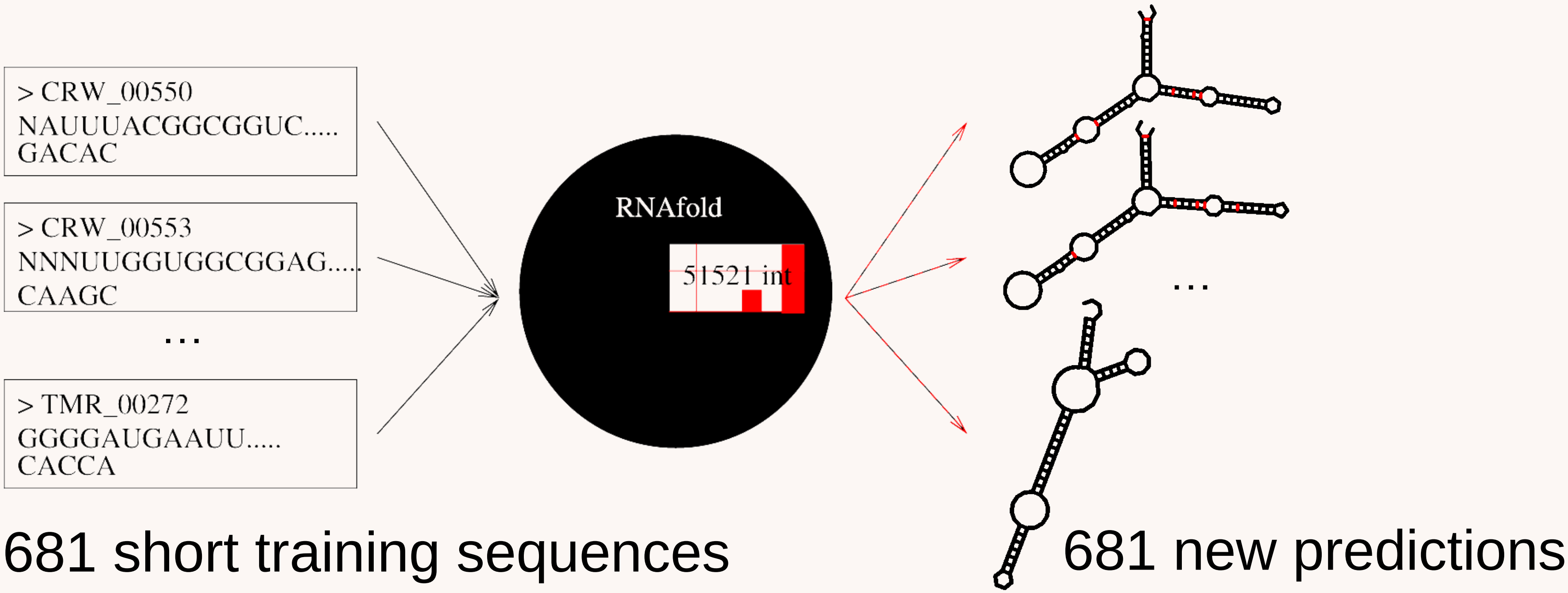


RNA sequence

RNA structure

RNAfold reads RNA molecules base sequence.
Outputs prediction of how molecule will fold up.
Internally RNAfold uses 51521 parameters.

Fitness of Mutated RNAfold



- Mutate constants inside RNAfold and recompile
- Run mutated RNAfold on training RNA sequences
- Compare each new prediction with real structure
- Fitness mean Matthew's correlation coefficient on 681 training RNA molecules

Genetic Improvement of RNAfold

RNAfold state of the art prediction of how RNA molecule will fold up based on its sequence of bases.

- Speed up via Intel SSE parallel instructions GI 2017. Shipped since V2.3.5
- ViennaRNA Package v2.3.0cuda
- Better predictions by evolving parameters
  - On average better predictions of RNA folding.
  - Shipped since 2.4.7
- AVX speedup will be in release 2.4.11

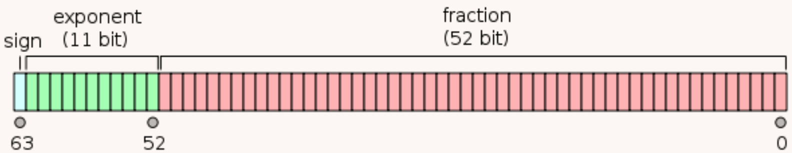
CREST

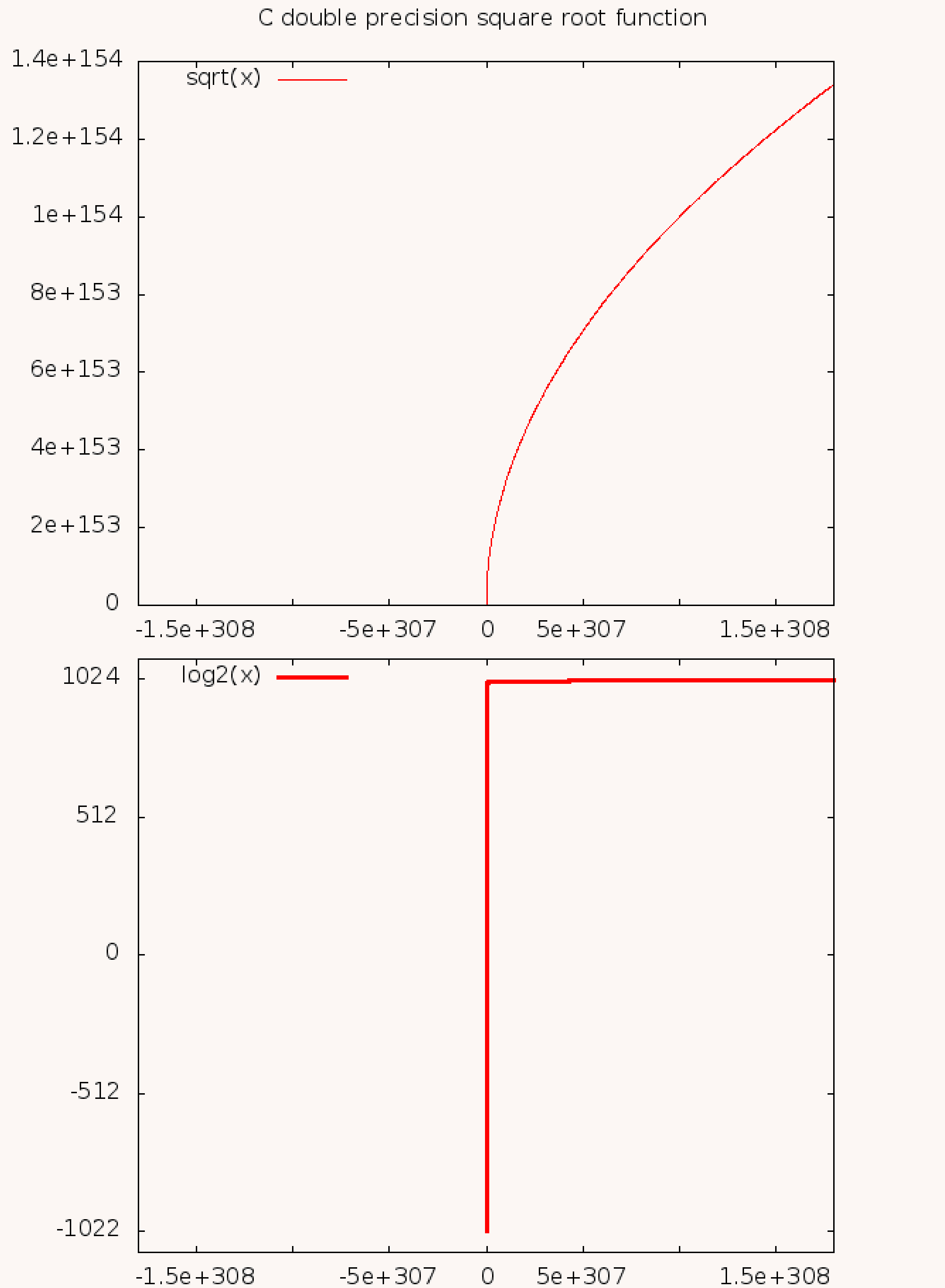
Use evolution to convert table based **sqrt into $\log_2$**

By updating table of 512 floats

(+small code change)

IEEE 754 Double Precision





GNU C library sqrt converted to log2

- Chosen implementation of sqrt divides normalised input into 512 bins.
- bin holds start point for Newton-Raphson
- 1 run evolutionary strategy per table value
  - Seed CMA-ES with square root values
    - Run code with CMA-ES generated table value
    - Fitness = absolute diff GI_$\log_2$(x) , log2(x)

      x takes three test values: smallest, mid, max in bin
  - all 512 CMA-ES runs succeed

Newton-Raphson is iterative way to find roots of continuous function using its derivative

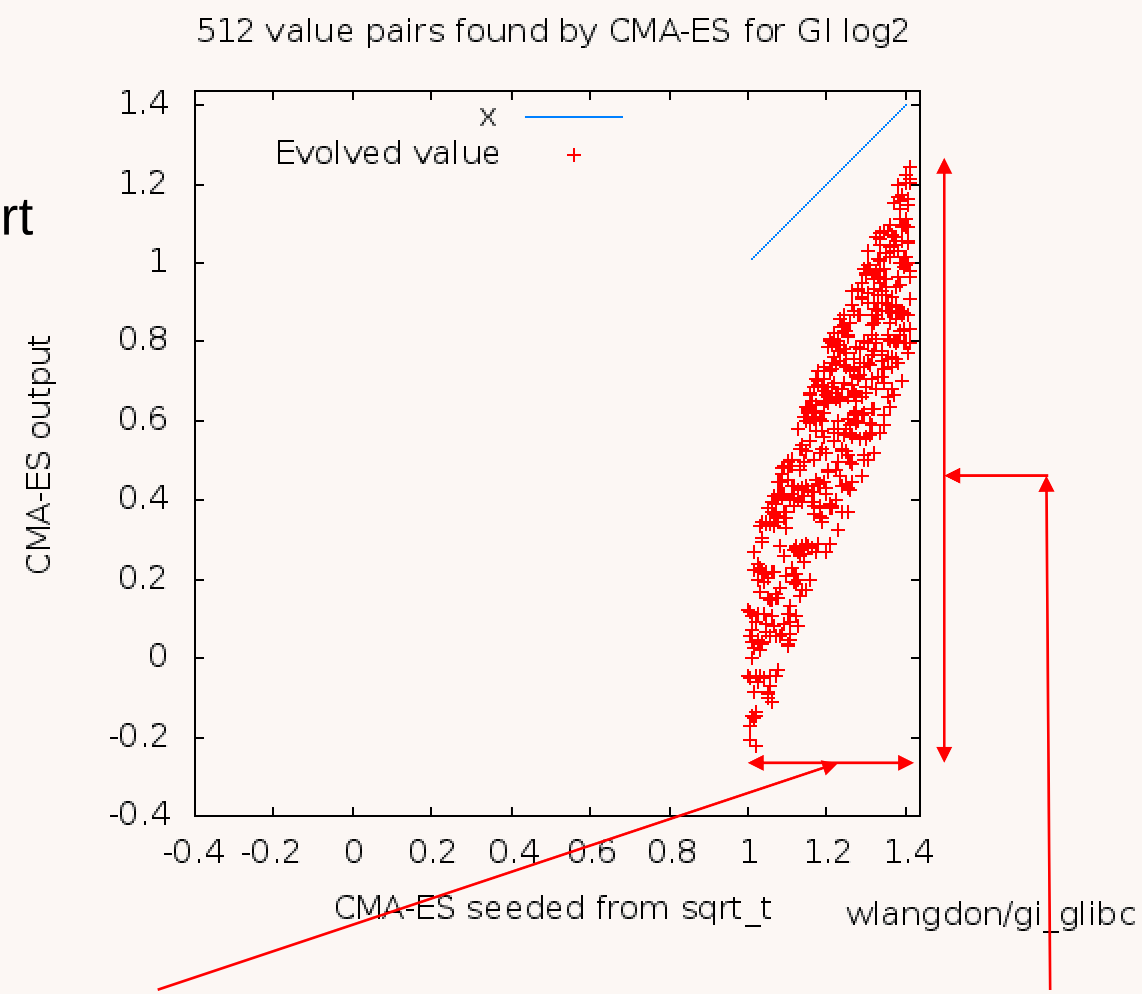
CREST

CMA-ES seeded with sqrt values

CMA-ES evolutionary run started with

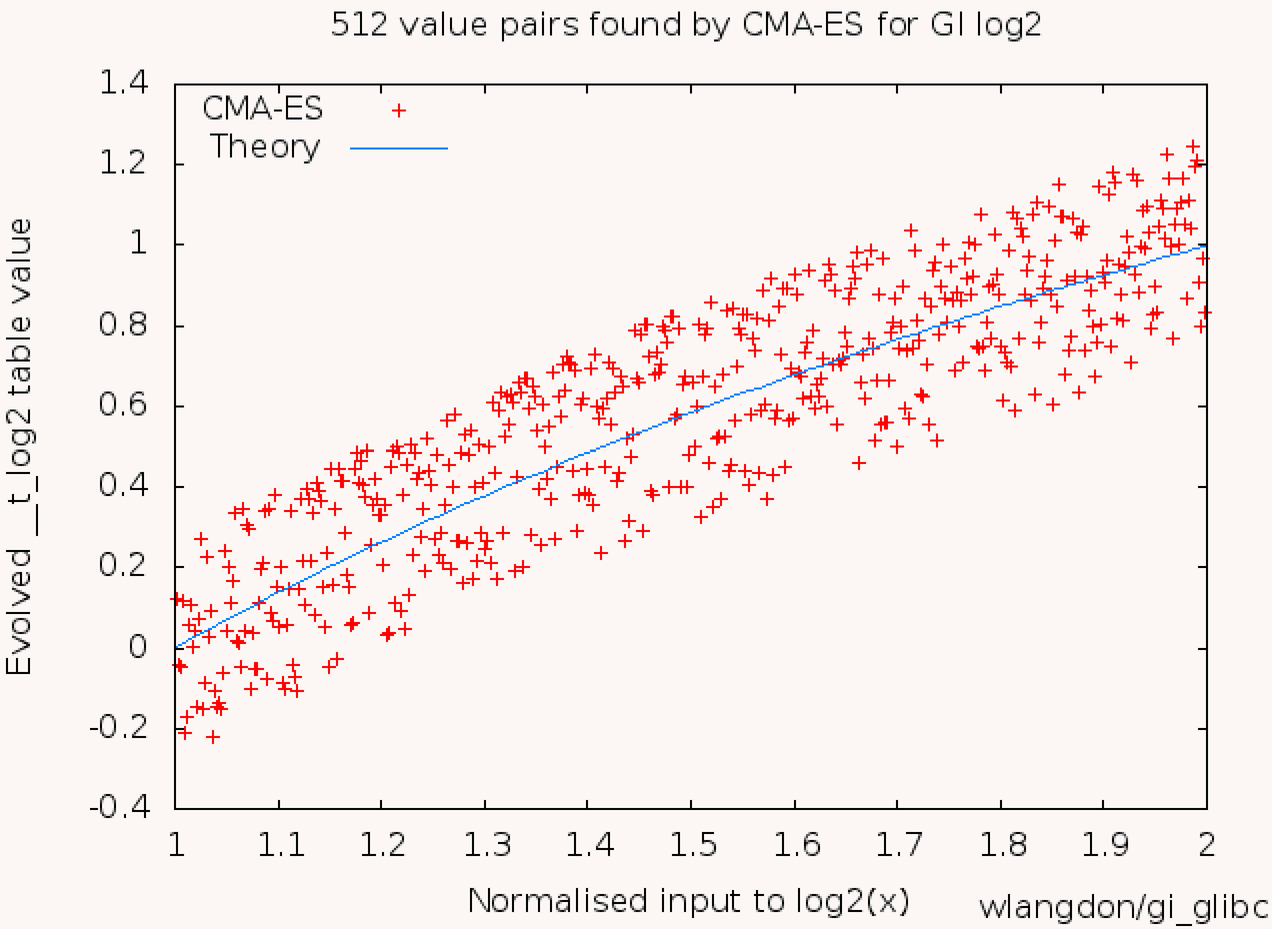
- value from sqrt
- mutation step size from variability of sqrt table

CMA-ES very easy to solve



Initial seed values are close to $\log_2$ values

Evolved values for Newton-Raphson



Evolved value scattered about theory

Evolved $\log_2$

New table driven function tested many thousands of times:

- Same precision as GNU C sqrt():
- Almost always exact
- Worst case last bit double precision $2.2\ 10^{-16}$

Automatic Software Maintenance

- In a world addicted to software, maintenance is the dominant cost of computing.
- Need to keep parameters up to date. Eg:
  - New science (cf. RNAfold), new laws or regulations, new users, new user expectations
  - Change of load, new hardware (eg bigger RAM), automatic porting
  - Search can be fast: cbrt < 5 minutes, log2 6secs
- Little SBSE research
- Great scope for automation

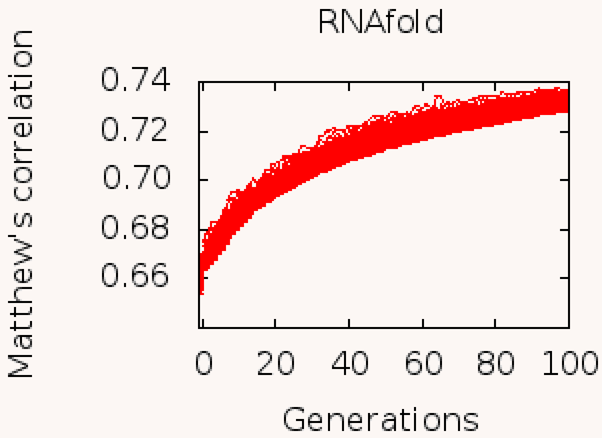
Six impossible things before breakfast

- To have impact do something considered impossible.
- If you believe software is fragile you will not only be wrong but shut out the possibility of mutating it into something better.
- Genetic Improvement has repeatedly shown mutation need not be disastrous and can lead to great things.

CREST

Summary

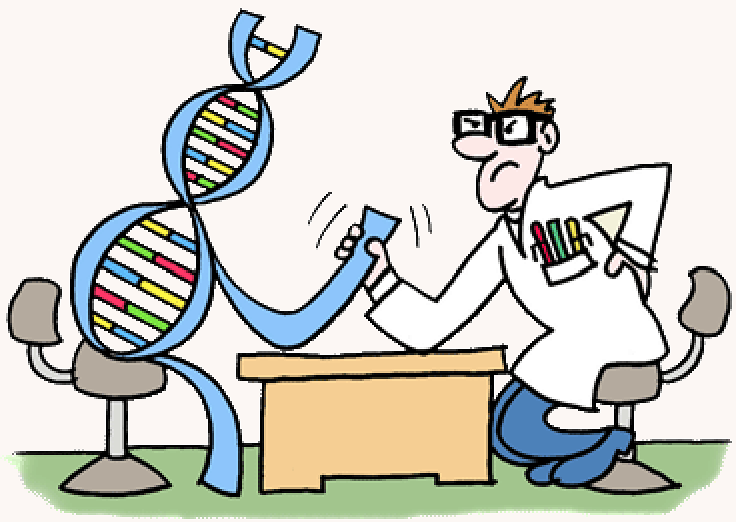
- Problem of maintaining data in code ignored
- SBSE to optimize data
  - suitable training data
  - treat code as a black box.
- RNAfold on real data
  - 50000 parameters 20% overall better prediction
  - AVX 45% speedup



- Rapidly generated new maths (cbrt, $\log_2$)
- Need research on Automatic Data Tuning
- Software is not fragile

GI 2019

GI 2019, Montreal, ICSE-2019 workshop.
Submission (2 or 8 pages) due 1st February 2019



Humies: Human-Competitive
Cash prizes
GECCO-2019

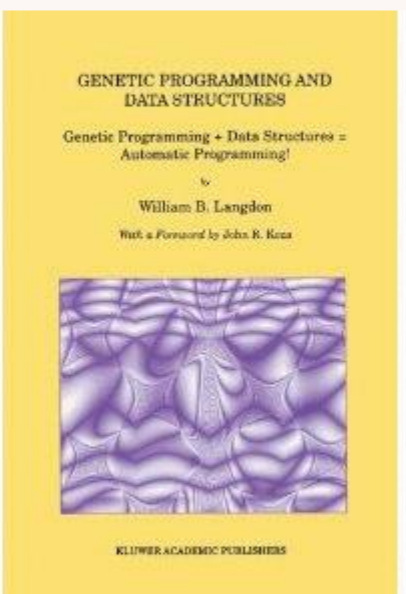

END OF DAASE

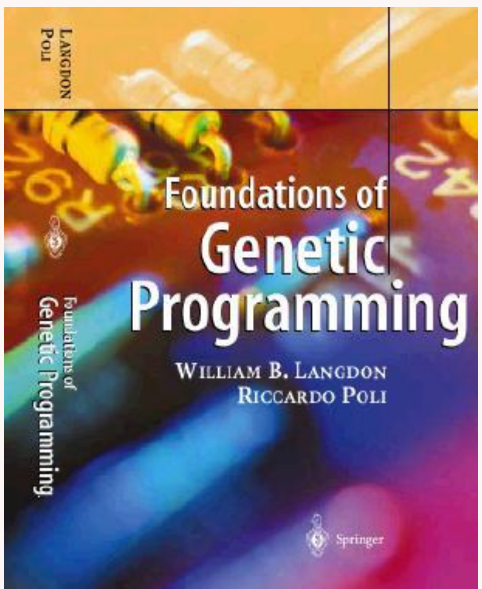
Genetic Improvement

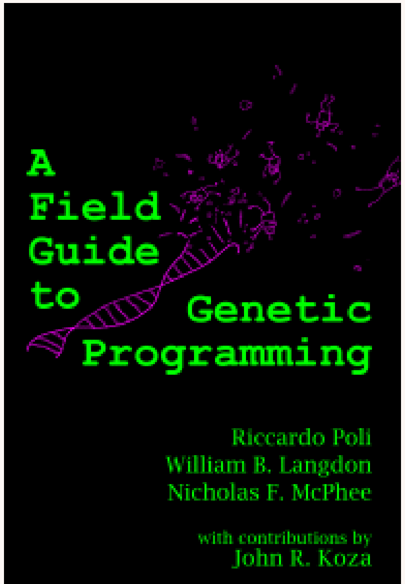
W. B. Langdon

CREST

Department of Computer Science







Improving RNAfold parameters

EuroGP-2018

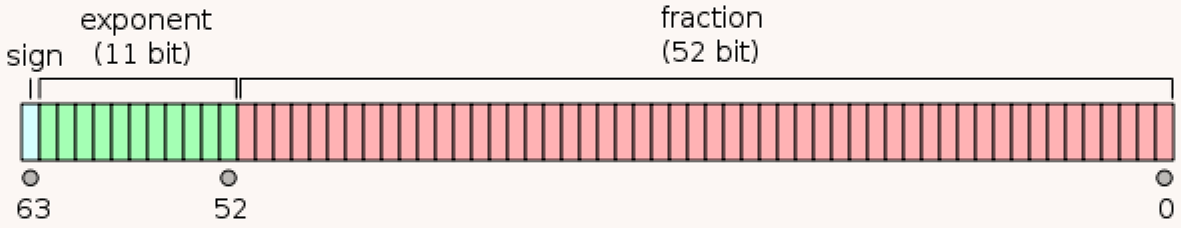
- RNAfold 7100 lines of C source code, 51521 parameters.
- Fitness correlation between prediction and true structure (MCC).
- Post evolution tidy
- 14732 (29%) parameters changed
- Holdout set significant increase in MCC
- Also better than constrained optimisation
- GI parameters rna_langdon2018.par shipped with ViennaRNA since 13 Jun 2018

Cube Root Code Changes I

- Most implementations of square root use hardware support.
- GNU C library glibc 2.27 also includes Newton-Raphson iterative solution
- Trap bad values, e.g. negative
- Normalise double input to 0.5 .. 2.0
- Guaranteed convergence in three steps:
  – Update both estimate of $\sqrt{x}$ and derivative
- Apply square root to exponent, ie divide by 2

Cube Root Code Changes II

- Remove trap for negative values
- Normalise double precision input to 1.0..2
  - Update both estimate of cube root $x^{1/3}$ and its derivative $\frac{1}{3}x^{-2/3}$
- Apply cube root to exponent, ie divide by 3



Evolved $\log_2$

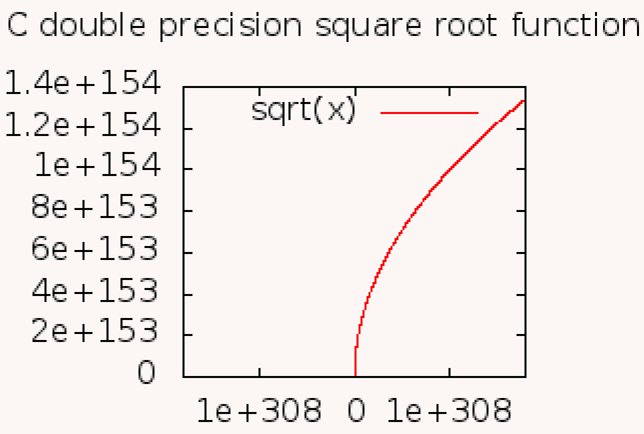
New table driven function tested many thousands of times:

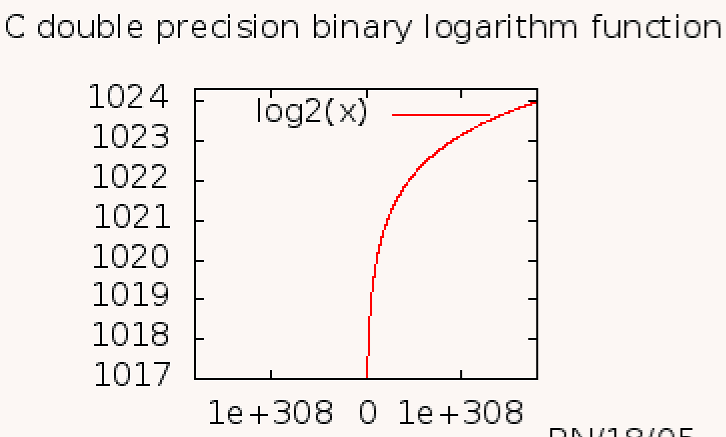
- Almost always exact (double precision).
- GI log2 always (except on NaN) returned a double y such that GNU C library exp2(y) is exactly x
  - or y differs only in least significant bit from the closest value which could be inverted by exp2 to yield x.
- Worst case last bit double precision 2.2 10^{-16}
- Same precision as GNU C sqrt()

Square root to binary log [RN/18/05](RN/18/05)

Frame work as sqrt to cbrt but

- Derivative known
- CMA-ES one dimension at a time (512 times) very easy





RN/18/05

The Genetic Programming Bibliography

http://www.cs.bham.ac.uk/~wbl/biblio/

12653 references, [11000 authors]()

Make sure it has all of your papers!

E.g. email W.Langdon@cs.ucl.ac.uk or use | [Add to It]() | web link

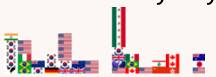
RSS Support available through the Collection of CS Bibliographies.

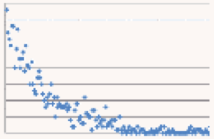
Downloads

A personalised list of every author's GP publications.

[blog]()

Downloads by day





Your papers

Search the GP Bibliography at

http://liinwww.ira.uka.de/bibliography/Ai/genetic.programming.html