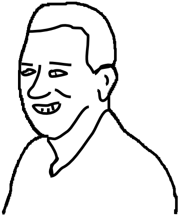
The SIAM 100.0000000-Digit Challenge: A Study in High Accuracy Numerical Computing Using Interval Analysis and *Mathematica*

James Rohal

The College of Wooster
Department of Mathematics

April 30, 2007



Advised By: Dr. Charles Hampton
Second Reader: Dr. Derek Newland

SIAM 100-Dollar 100-Digit Challenge

- Contest created by Lloyd N. Trefethen of Oxford University.
- Officially launched in the February 2002 Issue of *SIAM News*.
- 10 problems, 1 point per digit, maximum of 10 per problem.
- "Hint: They're hard! If anyone gets 50 digits in total, I will be impressed."

Interval Analysis

- Seldom used in practice.
    - Slowness of interval arithmetic packages.
    - Slow interval algorithms.
    - Difficulty of some interval problems.
    - Problems measuring time complexity.
- Benefits are enormous.
    - Help solve problems that noninterval methods cannot.
    - Guaranteed error bounds provide verifiably correct solutions.
    - More reliable since they usually converge.
    - Natural stopping criteria.

One Photon, Infinite Mirrors

Problem 1

A photon moving at speed 1 in the x-y plane starts at time $t = 0$ at $(x,y) = (1/2,1/10)$ heading due east. Around every integer lattice point (i,j) in the plane, a circular mirror of radius $1/3$ has been erected. How far from $(0,0)$ is the photon at $t = 10$?

Difficulties

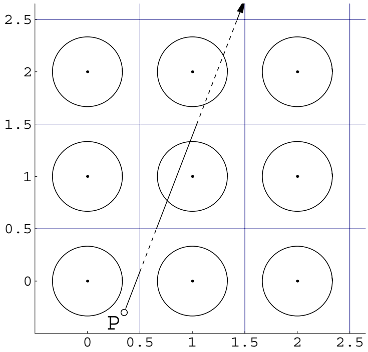
- This would be hard if this were not an ideal set up.
- Machine error enters quickly since 1/10 is not finitely representable in binary.
- Must have enough precision to guarantee that the reflected photon travels in the right direction!

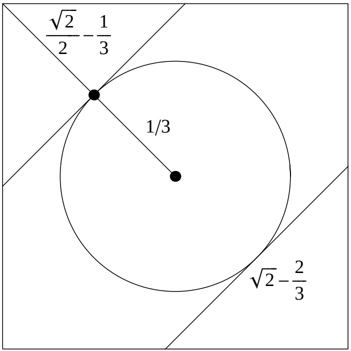
A Method of Approach

While t_{rem} is less than 10 do the following:

1. Find the next mirror of intersection.
2. Update the photon's position.
3. Update the photon's velocity.
4. Reduce the travel time of the photon from t_{rem}.

Find the Next Mirror of Intersection





Consider the sequence of mirrors corresponding to $P + (2/3)v$, $P + 2 \cdot (2/3)v$, $P + 3 \cdot (2/3)v, \ldots$ as long as necessary.

Update the Photon's Position and Velocity

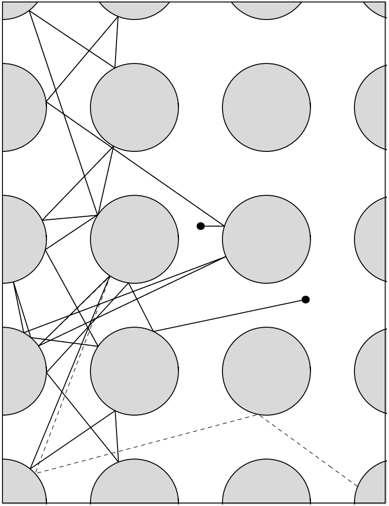
Let m be the center of the mirror corresponding to $P + k \cdot (2/3)v$, $k \in \mathbb{N}$.

1. $(P + tv - m) \cdot (P + tv - m) = 1/9$.
2. If s is the smallest positive root, then $Q = P + sv$.
3. H sends $(-a, -b)$ to (a, b), and fixes $(-b, a)$:

$$H \cdot \begin{pmatrix} -a & -b \\ -b & a \end{pmatrix} = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}.$$

But $(a, b) = Q - m$.

Results



A Naive Approach

- Put a small interval around each of our inputs and replacing each operation by its corresponding interval version.
- Stop running our algorithm if our interval does not have the required precision and restart algorithm with smaller interval enclosures.

Subtleties

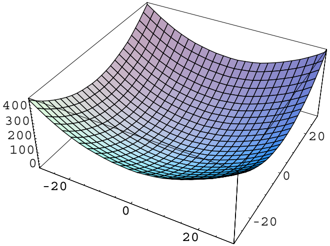
1. Use $s+2$ digits of working precision when the initial conditions have radius 10^{-s}.
2. Make sure none of the intermediate results have a precision less than the desired precision.
3. Check that the solution to our quadratic has no expression of the form $\sqrt{[\text{negative value}, \text{positive value}]}$.

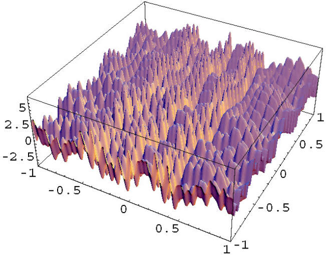
Hidden Complexity

Problem 2

What is the global minimum of the function

$$e^{\sin(50x)} + \sin\left(60e^{y}\right) + \sin(70\sin x) + \sin\left(\sin(80y)\right) - \sin\left(10(x+y)\right) + \frac{x^2+y^2}{4}?$$





Difficulties

- Can't solve $\left\{f_x = 0, f_y = 0\right\}$ easily since there are 2720 critical points.
- Look at the complexity near the origin! At first glance we can only estimate that the minimum lies within $[-1,1]\times[-1,1]$.

Grid Search

An easy way to get an upper bound on the minimum. This is sped up using compiled *Mathematica* expressions:

```
grid = Flatten[Table[{x, y}, {x, -1, 1, 0.01}, {y, -1, 1, 0.01}], 1];
fgrid = fcl /@ grid;
{Min[fgrid], Flatten[Extract[grid, Position[fgrid, Min[fgrid]]], 1]}

{-3.24646, {-0.02, 0.21}}
```

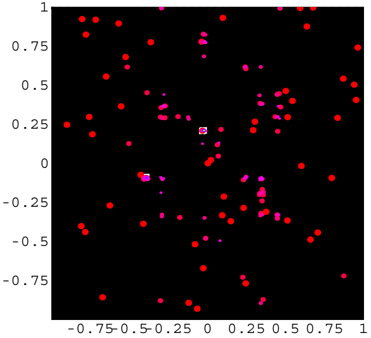
```
grid = Flatten[Table[{x, y}, {x, -1, 1, 0.001}, {y, -1, 1, 0.001}], 1];
fgrid = fcl /@ grid;
{Min[fgrid], Flatten[Extract[grid, Position[fgrid, Min[fgrid]]], 1]}

{-3.30563, {-0.024, 0.211}}
```

A Genetic Algorithm

- Inspired by biological evolution such as natural selection.
- Evolution starts from a random population.
- At each generation, the fitness of each individual is evaluated.
- If the fitness function yields a positive result, the individual lives.
- The remaining population is used in the next iteration.

Survival of the Fittest

- The points that survive each generation are called parents.
- The new points introduced at each generation are called children.
- Our fitness function evaluates each point and checks whether it is less than the upper bound.



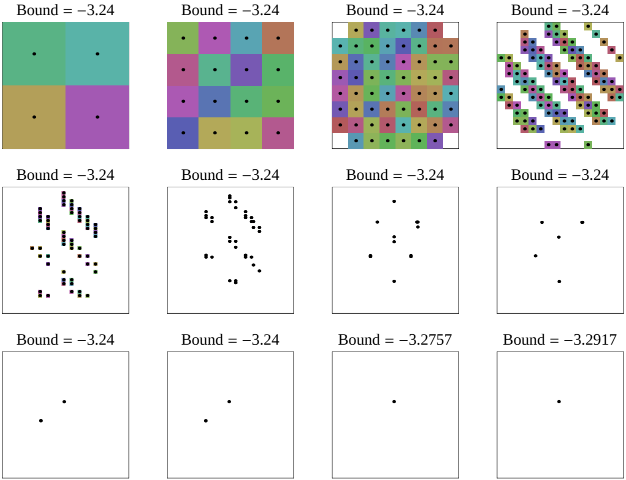
Search & Destroy

Subdivide R into smaller rectangles and retain only those rectangles T which pass the following conditions.
$f[T] = \{f(t) : t \in T\}$.

1. $f[T]$ is an interval whose left end is less than or equal to the current upper bound on the absolute minimum.
2. $f_x[T]$ is an interval whose left end is negative and right end is positive.
3. $f_y[T]$ is an interval whose left end is negative and right end is positive.

Search & Destroy

A tolerance $\varepsilon = 10^{-12}$ yields 12 digits after 47 iterations.



Newton Operator

Definition

If $f : \mathbb{R}^n \to \mathbb{R}^n$ is continuously differentiable and F is the interval extension of f, then the Newton operator on the n-dimensional box X is

$$N(X) := m(X) - J^{-1} \cdot F\big(m(X)\big),$$

where J is the interval Jacobian

$$F'(X) := \left(\frac{\partial F_i}{\partial x_j}(X)\right)_{ij} \quad \text{for } i,j = 1,2,\ldots,n.$$

Newton's Method

Start with a rectangle R and for each subrectangle X do the following:

- The Newton condition: check whether $N(X) \subseteq X$ holds. If so, then f has at most one zero in X.
- If $N(X) \cap X = \emptyset$, then there are no zeros in X.
- If neither situation applies, subdivide and try again.

If $N(X) \subseteq X$ holds, then let $X_1 := N(X) \cap X$ and repeat using X_1. Thus $X_n \subseteq X_{n-1} \subseteq \cdots \subseteq X_1 \subseteq X_0$.

Difficulties

- May converge to a real number in $N(X) \cap X$ rather than a root.
- Must approximate the interval value of J^{-1}.
- There may be unresolved roots at the end of our algorithm.
- There is no guarantee that $N(X) \subseteq X$ implies f has a zero in $N(X)$ for higher dimensions of f.

Krawczyk Operator

Definition

Suppose $f: D \subseteq \mathbb{R}^n \to \mathbb{R}^n$ is continuously differentiable in the open domain D and assume that f and f' have continuous interval extensions F and F' defined on interval vectors contained in D. Let $X = (X_1, X_2, \ldots, X_n)$ be a finite box contained in D where $X_1, X_2, \ldots, X_n$ are closed bounded real intervals. Then the Krawczyk operator is

$$K(X) := m(X) - YF\Big(m(X)\Big) + (I - YJ)\Big(X - m(X)\Big)$$

where J is the interval Jacobian, Y is the inverse of the matrix of midpoints of the intervals in J, and I is the $n \times n$ identity matrix.

Krawczyk Method

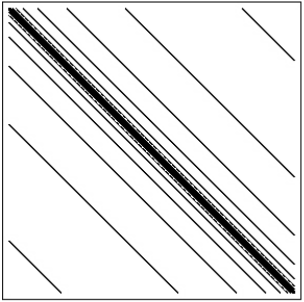
Let $\mathcal{R}$ denote the set of candidate rectangles.

- If $\mathcal{R}$ contains only one rectangle X, compute $K(X)$.
- If $K(X) \cap X = \emptyset$, there there is no critical point, and the minimum is on the border. So do nothing.
- If $K(X) \subset X$, then iterate the K operator starting with $K(X)$ until the desired tolerance is reached.
- Use the last rectangle to set the lower and upper bounds on the minimum and end the algorithm.

A Daunting Matrix

Problem 3

Let A be the $20,000 \times 20,000$ matrix whose entries are zero everywhere except for the primes $2, 3, 5, 7, \ldots, 224737$ along the main diagonal and the number 1 in all the positions a_{ij} with $|i - j| = 1, 2, 4, 8, \ldots, 16384$. What is the $(1,1)$ entry of A^{-1}?



Difficulties

- Inverting a matrix takes forever and is prone to error.
- Computing resources are limited, can't use techniques like Cramer's rule.
- Need to find properties that make this problem solvable!

The Approach

- $A\mathbf{x} = \mathbf{b}$ with $\mathbf{b} = (1, 0, 0, \ldots, 0)^T$ and $\mathbf{x} = (x_1, \ldots, x_n)^T$.

$$A^{-1} = \left(\begin{array}{c|ccc} x_1 & * & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ x_n & * & \cdots & * \end{array} \right)$$

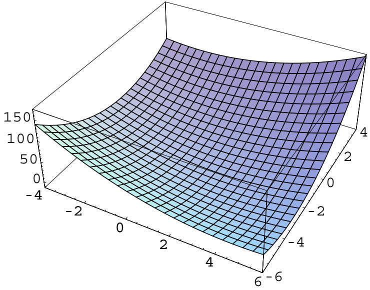
- What we want to find is the value of x_1.
- Symmetric.
- Positive-definite $\Rightarrow$ there exists a nonsingular matrix M such that $A = MM^T$.

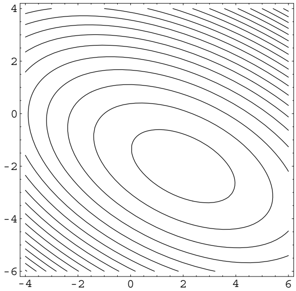
Quadratic Forms

Definition

If A is a matrix, $\mathbf{b}$ is a vector, and $c \in \mathbb{R}$ then the quadratic form is a scalar, quadratic function

$$f(\mathbf{x}) := \frac{1}{2}\mathbf{x}^T A\mathbf{x} - \mathbf{b}^T\mathbf{x} + c.$$





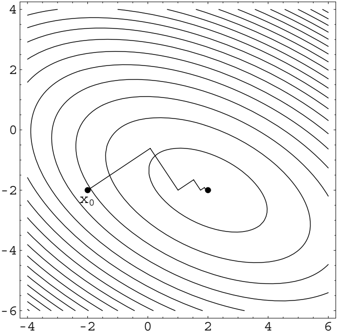
Quadratic Forms

Theorem

If A is a symmetric positive definite matrix, the solution to $A\mathbf{x} = \mathbf{b}$ is a critical point of $f(\mathbf{x})$. In fact, $\mathbf{x}$ is equal to the global minimum of $f(\mathbf{x})$.

Steepest Descent Method

Start at an arbitrary point $\mathbf{x}_0$ and slide down to the bottom of the paraboloid defined by $f(\mathbf{x})$. We take a series of steps $\mathbf{x}_1, \mathbf{x}_2, \ldots$ until we come within a reasonable distance from the true solution $\mathbf{x}$. Each step is proportional the negative of the gradient at $\mathbf{x}_i$.



The Method of Conjugate Directions

Choose the search directions so we don't take the same step more than once.

Definition

We say that two vectors $\mathbf{d}_i$ and $\mathbf{d}_j$, $i \neq j$, are *A-orthogonal* or *conjugate* if

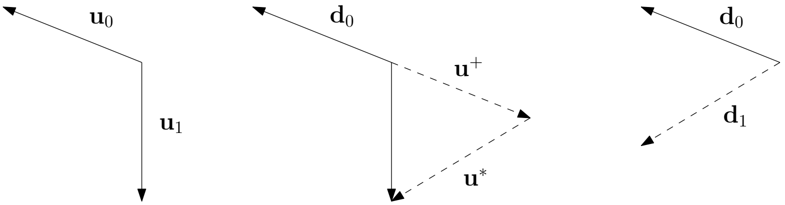
$$\mathbf{d}_i^T A \mathbf{d}_j = 0.$$

Theorem

The method of Conjugate Directions converges in n steps.

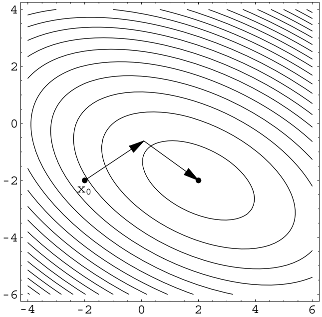
Conjugate Gram-Schmidt Process

Generate conjugate directions from a set of n linearly independent vectors $\mathbf{u}_0, \mathbf{u}_1, \ldots, \mathbf{u}_{n-1}$ by subtracting out any components that are not A-orthogonal.



Conjugate Gradient Method

- Let the set of linearly independent vectors be residuals: $\mathbf{r}_i = A - \mathbf{b}\mathbf{x}_i$.
- Reduces the time and space complexity from $O(n^2)$ to $O(m)$ where m is the number of non-zero entries in the matrix A.



Preconditioned Conjugate Gradient Method

Definition

The spectral condition number is

$$\kappa(A) = \lambda_{\max}(A)/\lambda_{\min}(A).$$

An ill-conditioned matrix is one in which the condition number is large.

- We want to find a matrix M such that $\kappa(M^{-1}A) \approx 1$. Thus we can apply the CG method to the system

$$M^{-1}A\mathbf{x} = M^{-1}\mathbf{b}.$$

- We use the matrix of the diagonal entries of A.