Cross Entropy Method and BFGS for Optimization of 2-Dimensional Lennard-Jones Clusters

Zaymon Foulds-Cook
Griffith University School of ICT

***Abstract*—Experimental findings here**

I. Introduction

A. Problem Statement

Linear polymers or chain molecules are molecules that exist in the natural world that are linear sequences of bonded atoms where each atom k within the molecule (excluding terminal atoms) is only bonded to atoms k_{k-1} and k_{k+1}. According to Valence-Shell Electron-Pair Repulsion Theory all atoms in a molecule interact with all other atoms regardless of whether the atoms are explicitly bound together through chemical bonds.

The energy of the interaction between two atoms is a function of the displacement between then and is given by the Lennard-Jones potential:

$$V = (\frac{1}{r^{12}} - \frac{2}{r^6})$$

where

$$r_{ij}^2 = x_{ij}^2 + y_{ij}^2$$

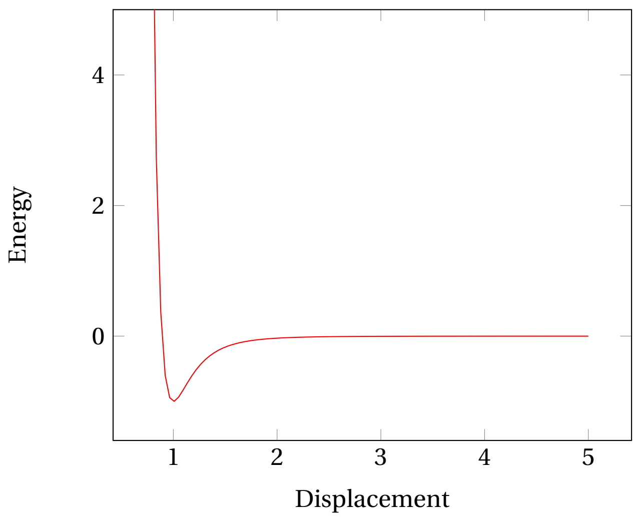

Figure 1. Energy as a function of displacement

The Lennard-Jones or "L-J potential" is simple mathematical approximation of the strength of interaction between a pair of neutrally charges atoms. As the distance between two neutrally charged similar atoms converges to 0 atoms experience Exchange Interaction or Pauli repulsion due to overlapping electron orbitals. The L-J Potential is an effective approximation commonly used as it reliably approximates energies at short and long distances. The L-J Potential is also used due to the computational simplicity since r_{ij}^{12} can be expressed as the square of r_{ij}^6.

For a molecule configuration to be stable it must be in a state of minimum energy. If the molecule is not in a configuration that results in minimum energy the configuration of that molecule would be transient in the physical sense as atom pairs throughout the molecule will be repelled and attracted. To calculate the total energy in a molecule configuration the summation of each atom pair's energy contribution is calculated. This is expressed by:

$$V = \sum_{i<j}^{N} (\frac{1}{r^{12}} - \frac{2}{r^6})$$

Designing computational methods for estimating and predicting molecule configurations has relevance to the study of: protein folding, molecular medicine, molecular physics, pharmacology and other fields examining the behavior and design of molecules and compounds.

B. Problem Representation

The problem can be represented by a vector of angles $\alpha_0 ... \alpha_{N-2}$ where each angle is relative to the previous angle and the angle's value ranges from $-\pi$ to π radians. This system is visualized in Figure 2. Each bond in the molecule is of length 1, ensuring the minimum energy between bonded molecules as demonstrated by the function plotted in Figure 1.

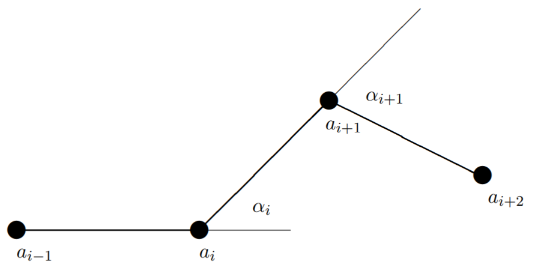

Figure 2. Problem configuration

There are several distinct advantages to this approach compared to using cartesian coordinates:

1) Specifying the angles removes the need to recalculate the positions of atoms after a change (as the angles are relative).
2) The cartesian coordinates can be obtained for any atom via trigonometry.
3) The vector of angles is a simpler representation of the problem and the change in energy for the cluster relative to the change in any α_i can be determined.

II. Literature Review

A. BFGS

The Broyden-Fletcher-Goldfarb-Shanno algorithm is an iterative method under the quasi-Newton class of hill climbing algorithms. "Quasi-Newton methods, like steepest descent, require only the gradient of the objective function to be supplied at each iteration" (Jorge Nocedal, 1999) The BFGS algorithm will never make a step in the 'uphill' direction. To effectively make use of BFGS in problem spaces with local minima BFGS needs to be used as the local optimization step in a higher level global optimization algorithm such as a genetic population based search.

B. Cross-Entropy Method

The Cross-Entropy method is a generic approach to solving complicated optimization problems such as Max Cut or the travelling salesman problem. "the CE method... defines a precise mathematical frame-work for deriving fast, and in some sense "optimal" updating/learning rules..." (Boer, Kroese, Mannor, & Rubinstein, 2005). The core algorithm of the CE method is quite a simple iterative process:

1) Generate a sample of random data according to some mechanism.
2) Score the sample and take an elite sub-sample of the resulting population.
3) Use the values of the elite sub sample to update the parameters of the random mechanism to produce a "better" sample on the next iteration.

The random mechanism can be as simple as specifying a random distribution for each element in the vector to be optimized (in this case the vector of angles) where the mean and standard deviation are changed to reflect the elite sample.

Specifying a smoothing parameter can be useful (Botev, Kroese, Rubinstein, & L'Ecuyer, n.d.). The smoothing parameter determines the ratio of blending between the old and new distribution parameters. This smoothing parameter λ can either be a static value $0 \le \lambda \le 1$ or dynamically changed as a function of time, score or another heuristic.

III. Algorithm Description

A. Cross Entropy Optimization

```
ALGORITHM CrossEntropy (sequenceLength):
    populationSize <- ALGORITHMPARAMETER
    parameterBlendRatio <- ALGORITHMPARAMETER
    eliteDistributionPercentage <- ALGORITHMPARAMETER
    epsilon <- ALGORITHMPARAMETER

    distributions <- [vector<normal distribution>] with default
          ↪ parameters
    population <- [vector<Sequence>] generatePopulation(
          ↪ populationSize, distributions)

    lastBest <- infinity
    bestScore <- infinity
    bestSequence <- Sequence

    while True:
        # Sort Population by Cost
        scoredPopulation <- [vector<pair<cost, Sequence>>] sort(
              ↪ population)
        diff <- lastBest - scorePopulation[0].cost
        lastBest <- scoredPopulation[0].cost

        if lastBest < bestScore:
            bestScore = lastBest

        # Update Distributions
        for angle in 0 <= angle < sequenceLength:
            values <- []
            for memberIndex in 0 <= memberIndex < populationSize *
                  ↪ eliteDistributionPercentage:
                values <+ scoredPopulation[memberIndex].Sequence[angle
                      ↪ ]

            mean <- mean(values)
            stdDeviation <- stdDeviation(values)

            distributions[angle].updateParameters(mean, stdDeviation)

        # Check exit conditions
        if (diff < epsilon):
            print bestScore
            break

        # Generate new Population
        population = generatePopulation(populationSize, distributions)
```

B. BFGS and Genetic Search

```
FUNCTION BFGS(sequence):
    gradientVector = sequence.CalcGradients


ALGORITHM GA_BFGS(sequenceLength):
    populationSize <- ALGORITHMPARAMETER
    epsilon <- ALGORITHMPARAMETER

    # Mutation chances
    mutation1 <- ALGORITHMPARAMETER
    mutation2 <- ALGORITHMPARAMETER
    mutation3 <- ALGORITHMPARAMETER

    population <- [vector<Sequence>] generatePopulation(
          ↪ sequenceLength)

    lastBest <- infinity
    bestScore <- infinity
    bestSequence <- Sequence

    while true:
        for member in population:
            member <- BFGS(member)

        scoredPopulation <- [vector<pair<cost, Sequence>>] sort(
              ↪ population)
        diff <- lastBest - scorePopulation[0].cost
        lastBest <- scoredPopulation[0].cost

        if lastBest <- bestScore:
            bestScore <- lastBest

        for member in population:
            if rand > mutation1:
                member.mutateRandomAngles()
                BFGS(member)
            if rand > mutation2:
                member.clearSubsequent()

        for randomPair in population:
            newPairs = crossOver(randomPair)
            randomPair.first = newPairs[0]
            randomPair.last = newPair[1]

        # Check exit conditions
        if (diff < epsilon):
            print bestScore
            break
```

C. Cross Entropy Method to Generate Optimization Candidates for BFGS

An experiment to use the Cross Entropy Method as the global optimizer in order to generate optimization candidates for BFGS. Conceptually this method will allow the Cross Entropy Method to optimize the distribution vector to generate good candidates for BFGS optimization.

1) High level algorithm:

1) Generate Population using distribution vector.
2) Run BFGS on population members, generating an additional optimized population.
3) Calculate the scores of the optimized population and sort.
4) Update the distribution vector based on the pre-optimized sequences which correspond to the sequences in the elite sample of the optimized population.
5) Goto step 1.

IV. Algorithm Performance

A. CE Method

1) Tuning CE Parameters:

B. CE - Scores

N	Optimal	Best Found	N	Optimal	Best Found
4	-5.1	-5.07132	30	-77.2	-70.2578
5	-7.2	-7.15466	31	-79.5	-69.2296
6	-9.3	-9.29635	32	-82.8	-67.8379
7	-12.5	-12.427	33	-86.1	-75.3775
8	-14.7	-14.5394	34	-88.3	-77.1285
9	-16.9	-16.7548	35	-91.7	-81.6391
10	-20.1	-19.8421	36	-95.0	-82.3254
11	-22.3	-22.1631	37	-98.3	-89.7201
12	-25.5	-24.322	38	-100.5	-87.3879
13	-27.8	-27.5986	39	-103.8	-92.6629
14	-31.0	-29.7009	40	-107.1	-86.8567
15	-33.2	-32.8739	41	-109.4	-90.8365
16	-36.5	-33.1633	42	-112.7	-95.9652
17	-38.7	-35.2322	43	-116.0	-104.069
18	-42.0	-39.5358	44	-119.3	-102.404
19	-45.3	-42.1855	45	-121.6	-102.907
20	-47.5	-44.6016	46	-124.9	-104.417
21	-50.8	-47.1159	47	-128.6	-107.945
22	-53.0	-51.4069	48	-131.5	-109.423
23	-56.3	-48.5068	49	-133.8	-110.755
24	-59.6	-52.3969	50	-137.1	-114.764
25	-61.8	-53.2786	51	-140.5	-115.81
26	-65.1	-52.1118	52	-143.7	-116.132
27	-68.4	-61.5712	53	-146.0	-127.258
28	-70.0	-63.0996	54	-149.3	-124.026
29	-74.0	-64.3941	55	-152.7	-130.41

Table I
BEST SCORES FROM CROSS ENTROPY METHOD

C. CE Method - BFGS

N	Optimal	Best Found	N	Optimal	Best Found
4	-5.1	-5.0717	30	-77.2	-64.042
5	-7.2	-7.16219	31	-79.5	-66.6319
6	-9.3	-9.34017	32	-82.8	-70.008
7	-12.5	-12.465	33	-86.1	-70.969
8	-14.7	-14.6329	34	-88.3	-72.8008
9	-16.9	-16.6597	35	-91.7	-78.7509
10	-20.1	-18.9354	36	-95.0	-77.1246
11	-22.3	-21.4258	37	-98.3	-80.0404
12	-25.5	-24.3834	38	-100.5	-79.7988
13	-27.8	-26.918	39	-103.8	-85.7043
14	-31.0	-29.6838	40	-107.1	-85.4764
15	-33.2	-32.6135	41	-109.4	-88.7995
16	-36.5	-33.9053	42	-112.7	-91.8655
17	-38.7	-36.9939	43	-116.0	-88.4889
18	-42.0	-38.1547	44	-119.3	-88.7862
19	-45.3	-41.6245	45	-121.6	-91.8972
20	-47.5	-44.8945	46	-124.9	-99.764
21	-50.8	-46.4786	47	-128.6	-98.9701
22	-53.0	-48.0152	48	-131.5	-102.948
23	-56.3	-51	49	-133.8	-97.7711
24	-59.6	-52.6622	50	-137.1	-103.551
25	-61.8	-54.948	51	-140.5	-105.572
26	-65.1	-57.8007	52	-143.7	-107.815
27	-68.4	-56.8625	53	-146.0	-106.928
28	-70.0	-60.2806	54	-149.3	-109.118
29	-74.0	-62.7181	55	-152.7	-118.146

Table II
BEST SCORES FROM COMBINATION OF BFGS AND CROSS ENTROPY METHOD

References

Boer, P.-T. B., Kroese, D. P., Mannor, S., & Rubinstein, R. Y. (2005). A tutorial on the cross-entropy method. *Annals of Operations Research*, *134*, 19–67. Retrieved from `https://people.smp.uq.edu.au/DirkKroese/ps/aortut.pdf`

Botev, Z. I., Kroese, D. P., Rubinstein, R. Y., & L'Ecuyer, P. (n.d.). *The cross-entropy method for optimization.* Retrieved from `https://people.smp.uq.edu.au/DirkKroese/ps/CEopt.pdf`

Jorge Nocedal, S. J. W. (1999). *Numerical optimization.* New York: Springer-Verlag. Retrieved from `http://www.bioinfo.org.cn/ wangchao/maa/Numerical_Optimization.pdf`

Appendix A
Gradient Calculation

The total energy of the system is given by:

$$V = \sum_{i<j}^{N} (\frac{1}{r_{ij}^{12}} - \frac{2}{r_{ij}^{6}})$$

where

$$r_{ij}^2 = x_{ij}^2 + y_{ij}^2$$

Let

$$\Psi_k = \sum_{i-1}^{k} \alpha_i$$

and $(x_0, y_0) = (0,0)$ and $\Psi_0 = 0$ then

$$x_i = \sum_{k=0}^{i-1} \cos \Psi_k$$

$$y_i = \sum_{k=0}^{i-1} \sin \Psi_k$$

Now

$$V_{\alpha m} = \frac{-12 \sum_{i<j} (\frac{1}{r_{ij}^{13}} - \frac{1}{r_{ij}^{7}})}{r_{ij}}$$

and

$$(r_{ij})_{\alpha m} = \frac{((x_i - x_j)(x_i - x_j)_{\alpha m} + (y_i - y_j)(y_i - y_j)_{\alpha m}}{r_{ij}}$$

assuming that $j > i$ we have

$$(x_i - x_j)_{\alpha m} = -\sum_{k=i}^{j-1} (\cos \Psi_k)_{\alpha m} = \sum_{k=max(i,m)}^{j-1} \sin \Psi_k$$

$$(y_i - y_j)_{\alpha m} = -\sum_{k=i}^{j-1} (\sin \Psi_k)_{\alpha m} = \sum_{k=max(i,m)}^{j-1} \cos \Psi_k$$

Therefore

$$(r_{ij})_{\alpha m} = ((x_i - x_j)(\sum_{k=max(i,m)}^{j-1} \sin \Psi_k) +$$

$$(y_i - y_j)(- \sum_{k=max(i,m)}^{j-1} \cos \Psi_k))/r_{ij}$$

which is zero when $m <= i$. Assuming $m > i$

$$-\sum_{k=m}^{j-1} \cos \Psi_k = x_j - x_m$$

$$-\sum_{k=m}^{j-1} \sin \Psi_k = y_m - y_j$$

and

$$(r_{ij})_{\alpha m} = \frac{((x_i - x_j)(y_m - y_j) + (y_i - y_j)(x_j - x_m)_{\alpha m}}{r_{ij}}$$

Combining the terms we have

$$\frac{\partial V}{\partial \alpha m} = -12 \sum_{i=0}^{m-1} \sum_{j=m+1}^{n} (\frac{1}{r_{ij}^{14}} - \frac{1}{r_{ij}^{8}})((x_i - x_j)(y_m - y_i) + (y_i - y_j)(x_j - x_m))$$