Energy minimization of protein structure using homology modeling and Particle Swarm Optimization with Dynamic Inertia Weights

Yogendra Singh Kushwah and R K Shrivastava
Department of Mathematics, Government Science College, Gwalior, MP., India.

Abstract: Energy minimization in Bio-molecular system is an intensive application in Protein Structure Prediction (PSP). Traditional methods for reducing energy such as The Conjugate gradient method and Steepest descent method converge only local energy minima which are highly reliant on set the parameter for calculation in initial stage. In this paper we put forward a modified PSO (DIWPSO) with control parameter D which choose best Inertia Weight strategies for energy minimization in homologous protein structure, among four best Inertia weight strategies. The proposed approach (DIWPSO) significantly improvement to search deeper energy minima and find out global energy minimum. It is found that in result analysis, proposed algorithm gives better result with regard to rate of convergence and get to minimum energy compare to traditional methods. Our homology based DIWPSO check approach utilize the design appearing to make the half of the basic masses and staying half are made haphazardly and especially handle sub-atomic compliances with a sub-atomic part reenactment bundle (AMBER12).

Keywords: Particle swarm optimization, Inertia weight strategies, Protein Structure, Energy Minimization, Hybridization, Convergence.

I. Introduction

Inertia weight is an crucial parameter of PSO. It has the capacity to maintain to balance between exploration and exploitation. Inertia weight is used to provide the optimum velocity to the particle at the present time stamp. When first version of PSO presented by Eberhart and Kennedy in 1995 [1], then there was no concept of Inertia Weight given by him. First time, Inertia Weight was introduced by Shi and Eberhart [2] in 1998, with constant inertia weight. They proposed to higher and small Inertia Weight which helps in global and local search. Further there are many Inertia Weight strategies proposed by researchers that give better performance of PSO for different type of Optimization problems. A brief review of Inertia Weight strategies that is used in PSO are explained in subsequent Sections.

Bansal et al. [3] did work with different strategies of Inertia Weight. The main purpose of this studies that selecting the best inertia weight strategies. Table I shown different inertia weight strategies which are required in DIWPSO. In this paper propose a Dynamic Inertia Weight Particle Swarm Optimization Algorithm (DIWPSO) with control parameter D that choose inertia weight among four inertia weight strategies with equal probability to search those parameter of function which minimize the energy in bio-molecular system. rate of change of particle position can be updated through control parameter D which choose inertia weight from four inertia weight strategies on basis of equal probability. DIWPSO is tested over 25 linear, nonlinear, single variable, multivariable, Constrained and Non Constrained real world optimization problems. The results analysis shows that the modified PSO is better than basic PSO with respect to accuracy, rate of convergence and robustness.

Proteins perform all short of biological functions. These function is result of proper folded conformation adopt by proteins. Protein is a long polymer chain made by common set of 20 amino acids, joined by a covalent peptide bond. These protein are encode by gene in the cell system. 3D conformation prediction of protein form is primary linear sequence is important in understand the biological function. The fold protein has different energy level and lower energy state fold are more stable in nature, so all protein prediction methods consider energy minimization algorithm is important for the computation protein structure prediction.

Theoretical simulations for Protein folding have lot of significant problems. It is very difficult to search the sequence of protein’s amino acids in three dimensional atomic structures. It is also challenging problem that the long structure of protein molecule that has to be searched and energy minimization had to be covered. [4]. Anfinsen’s hypothesis also suggested that thermodynamically structure of protein might be predicted from the sequence of amino acids by minimizing an appropriate free energy function. Even in laboratory experiments

DOI: 10.9790/0661-1906042431 www.iosrjournals.org
correct folded protein is based on the minimum of total free energy has seen. Another significant problem is numerical analysis of energy optimization in energy function over native protein structures that calculate the global energy minimum has been difficult to define [5]. Therefore, this field attracted researcher to developing and optimizing simplified energy functions through parametrization [4], [6], [7], [8].

There are two Optimization method used for energy minimization. Threading [7], [9], [10], [11], [5] and the lattice models of folding [4], [12]. Another method is the decay-based parameterization [8], in which set the parameter that search the maximizing energy gap between decoy protein structure and the native protein structure. These methods are applicable for only when the protein structure is distinct. In recently free the above condition by the method [6], this method based on the assumption that different molecule reach to closer in native structure by set the parameter to optimize the energy function.

Above approaches to setting the parameters for energy minimization are based on experience. This parameter is chosen from favorable set of constraints for the function of energy minimization. There is no intensive work done in the field of setting the parameter to obtain the minimum energy. For this purpose propose a Dynamic Inertia Weight Particle Swarm Optimization Algorithm (DIWPSO) with control parameter D that choose inertia weight among four inertia weight strategies with equal probability to search those parameter of function which minimize the energy in bio-molecular system. DIWPSO technique reduces the hindrance to search local energy minima and find out deeper energy minima. This algorithm help to get lower energy found the global energy minima. With the help of AMBER12, molecular dynamics simulation package [13] we calculate the molecular energy of molecules during minimization.

Rest of the paper put in the following manner. In Section II describe about PSO. Various Inertia Weight Strategies for Particle Swarm Optimization are put in Section III. Implementation of DIWPSO explained in Section IV. Section V discusses the energy minimization using DIWPSO. Section VI refers to discussion about experiments and results. Finally, Section VII covered Conclusion.

### II. Particle Swarm Optimization

Particle Swarm Optimization is a swarm intelligence-based optimization technique. Kennedy and Eberhart [1] was given this technique in 1995. This technique based on collective social response in a group which shown in some Insects or Animals. Such type of behavior can be observe in fish schooling and bird flocking. In PSO, every individual is known as particle that is involved in the searching of food source (optimal solution). Every particle has the capability to remember own pbest solution which gained during searching the gbest solution. Every, \( i^{th} \) particle is described as \( y_i = (y_{i1}, y_{i2}, \ldots, y_{in}) \). Randomly generate the population is an initial solutions which search optimal solution. The pbest is the personal best position of an individual in the swarm. The gbest is the global best position among all the individual in the swarm. The \( i^{th} \) individual best position is represented as \( p_i = (p_{i1}, p_{i2}, \ldots, p_{in}) \). The velocity of a particle is the rate of change in the position and represented as \( v_i = (v_{i1}, v_{i2}, \ldots, v_{in}) \). The particles are update the velocity and position through following equations:

\[
\begin{align*}
\text{v}_d &= \text{w} \times \text{v}_d + c_1 \times r_1 \times (\text{p}_d - \text{y}_d) + c_2 \times r_2 \times (\text{p}_d - \text{y}_d) \\
\text{y}_d &= \text{y}_d + \text{v}_d
\end{align*}
\]

(1)

(2)

where \( w, c_1, c_2, r_1 \) and \( r_2 \) are parameters. \( d = \{1, 2, 3, \ldots, N\} \), \( w \) represents the inertia weight which provides equivalence in the activity of exploration and exploitation in the search space. \( c_1, c_2 \) are acceleration constants which enforce to each individual in the swarm to attain personal best and global best position. \( r_1 \) and \( r_2 \) represent random number which get the value in the interval \([0, 1] \). If put the large value of \( r_1 \) and \( r_2 \) then particle search speed is fast, in this situation particle can miss the gbest position. Small value of \( r_1 \) and \( r_2 \) breakdown the speed of the particle and may be he gets premature solution. \( v \) is the velocity of the individual which lies between \( v_{\text{min}} \) and \( v_{\text{max}} \). The range of the velocity is constrained of the individual in the solution space. Very small velocity might miss the opportunity of the particle to search global best solution and very high velocity might be unable to reach Optimal solution. The pseudo code of PSO algorithm is given in Algorithm (1).

**Algorithm 1:** Basic Particle Swarm Optimization

**Input:**
- Initialize the parameters
**Output:**
- (gbest ) molecule
**Begin:**
- for molecule do
  - introduce arbitrarily.
end for
iter=0
while iter < Maxiter OR Minerr condition not satisfied do
    for each molecule do
        Calculate wellness esteem.
        if new wellness esteem is superior than pbest then
            pbest = new calculated wellness esteem
        end if
    end for
    Find gbest molecule from the pbest molecules.
    for each molecule do
        Calculate molecule speed indicated by (1).
        Update molecule position concurring (2).
    end for
    iter = iter + 1
end while

III. Strategies Of Inertia Weight For Particle Swarm Optimization

Inertia weight is an important parameter of PSO. It has the capacity to maintain the balance between exploration and exploitation. Inertia weight is used to provide the optimum velocity to the particle at the present time stamp. When first version of PSO presented by Eberhart and Kennedy in 1995 [1], then there was no concept of Inertia Weight given by him. First time of Inertia Weight introduced by Shi and Eberhart [2] in 1998, with constant inertia weight. They proposed to higher and small Inertia Weight which helps in global and local search. Further there are many Inertia Weight strategies proposed by researchers that give better performance of PSO for different type of Optimization problems. A brief review of Inertia Weight strategies that is used in PSO are explained in subsequent Sections.

Bansal et al. [3] did work with different strategies of Inertia Weight. The main purpose of this study is that selecting the best inertia weight strategies. Table I shows different inertia weight strategies which are required in DIWPSO. In this paper rate of change of particle position can be updated through control parameter D which choose inertia weight from four inertia weight strategies on basis of equal probability. DIWPSO is tested over 25 linear, nonlinear, single variable, multivariable, constrained and non constrained optimization test problems taken from literature. The results analysis shows that the modified PSO is better than basic PSO with respect to accuracy, rate of convergence and robustness. Chen et al. [14] proposed two different Inertia Weight strategies which based on Decreasing Inertia Weight. It is also analysed that proposed strategies is better than with respect to convergence of solution in early stage of approximation. It is also found that proposed strategies improve the result for continuous optimization problem.

<table>
<thead>
<tr>
<th>SN</th>
<th>Inertia Weight strategies</th>
<th>Formula of Inertia Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Constant Inertia Weight [2]</td>
<td>( w = c )</td>
</tr>
<tr>
<td>2.</td>
<td>Random Inertia Weight [10]</td>
<td>( w = 0.5 + \text{rand}() / 2 )</td>
</tr>
<tr>
<td>3.</td>
<td>Chaotic Inertia Weight [6]</td>
<td>( z = 4 + z \cdot (1 - z) ) ( \frac{\text{MAX}_w - \text{ITER}}{\text{MAX}_w - \text{ITER}} ) + ( w ) ( z )</td>
</tr>
<tr>
<td>4.</td>
<td>Linear decreasing Inertia Weight [11]</td>
<td>( w_k = w_{\text{max}} - \frac{w_{\text{min}} - w_{\text{max}}}{\text{ITER}_{\text{max}}} \cdot k )</td>
</tr>
</tbody>
</table>

Fayek et al. [5] hybridize the Particle Swarm with Simulated Annealing (PSOSA) in which Inertial Weight was optimized and it was tested on urban planning problem and it found that the hybrid approach is better in respect to fast rate of convergence. It is also adaptive in urban planning problem to increased load of growing number of blocks. Chaotic Inertia Weight was proposed by Feng et al. [16]. It uses the merits of chaotic optimization. It is found that CRIW PSO performs better that RIWPSO. Malik et al. [17] proposed Sigmoid Increasing Inertia Weight strategy (SIIW) which has capability to fast convergence when fitness value of the function is minimum. In this approach Linear Increasing Inertia Weight combine with sigmoid function. To increase the overall convergence speed Gao et al. [18] proposed (LDIW), this is The Chaos mutation operator which was combined with Logarithm Decreasing Inertia Weight. The Chaos mutation increase the ability to come out of the local minima while LDW improves its convergence speed. Exponent Decreasing Inertia Weight (EDIW) was also proposed by Gao et al. [18] that overcome premature convergence. Also stochastic mutation [19] with Exponent Decreasing Inertia Weight improves the performance of PSO and getting out from local minima.
IV. DiwpsO: Dynamic Inertia Weights Based Particle Swarm Optimization

Bansal et al. [3] did work with different strategies of Inertia Weight. The main purpose of this study is that selecting the best inertia weight strategies. Table I shows different inertia weight strategies which are required in DIWPSO. In this paper, Dynamic Inertia Weight Particle Swarm Optimization Algorithm (DIWPSO) with control parameter D that choose inertia weight among four inertia weight strategies with equal probability to search those parameter of function which minimize the energy in bio-molecular system. rate of change of particle position can be updated through control parameter D which choose inertia weight from four inertia weight strategies on basis of equal probability. In the modified algorithm, the target is that the updating of molecule speed using selected inertia weight strategies presented in Table I. These strategies based on equal probability to update the molecule speed in compliance with control parameter D. The control parameter D has range in [0,1]. The algorithm (2) is the pseudo of DIWPSO.

Algorithm 2: Dynamic Inertia Weight Particle Swarm Optimization (DIWPSO)

Input:
Initialize the parameters
DIW* = [CoIW, RIW, ChIW, LDIW]
*ChIW = Chaotic Inertia Weight; RIW = Random Inertia
Weight; CoIW = Constant Inertia Weight;
LDIW = Linear decreasing Inertia Weight
Output:
\( (gbest) \) molecule

Begin:
for individual do
  initialize the swarm.
end for
iter = 0
while iter < Maxiter OR Minerr condition not satisfied do
  for each molecule do
    Calculate wellness esteem.
    if new wellness esteem is superior than pbest then
      pbest = new computed wellness esteem
    end if
  end for
  find gbest molecule from all pbest molecule.
  for each molecule do
    Choose inertia weight \( (w) \) using following strategy.
    If \( \text{rand}(i) \leq D \) then
      \( i = \text{uniformly distributed random Integer in} \ (1,4) \backslash \backslash \)
      \( w = \text{DIW} [i] \)
    end if
    Evaluate molecule speed as per (1).
    Update molecule position agreeing (2).
  end for
  iter = iter + 1
end while

V. Energy Minimization Using DiwpsO

The DIWPSO guides the demand in an educated way: exceptional parameters (to the degree satisfying lower noteworthiness scarcest) are held and misused to the most ludicrous degree through time, while new scopes of the parameter space are investigated intentionally through blend and change. The focal accepted is to encode the preoccupation parameters and conditions into strings, and apply the DIWPSO number to the strings with a target work mirroring the essentialness of the structure noteworthiness. We apply DIWPSO with a specific extreme goal to look for parameters that purpose of control the free vitality of a bio-sub-atomic framework. A subset of these parameters is generally chosen for progress and is encoded into DIWPSO strings. The strategy of parameters to be advanced is issue subordinate, and is picked in context of the physical prerequisites and setups of the structure. It is precious watching that a solitary vitality minimization figuring in AMBER12 has five particular stages: The hidden four phases are atomic segments eras, and the last stage is the
noteworthiness minimization step utilizing steepest jump or conjugate inclination technique. AMBER12 parameters contain a few classes concerning limit conditions, requirements, potential vitality limits, purpose of merging of mass improvements, non-reinforced affiliations, and program control parameters for these calculations. These five phases reflect a technique for instatement, warming, dependable temperature sub-atomic stream spread, cooling, and vitality minimization. We utilize the AMBER12 bundle [13] to enlist the vitality of bio-atomic structures.

Two delineation methods that are accessible for the bio sub atomic framework, i.e. In the present study we utilized true blue with respect encoding course of action for the parameters. Space settled Cartesian direction [22] and inside headings framework. A significant piece of the reviews have demonstrated that space-settled Cartesian heading delineation enhances the feasibility of the consider separated and interior supports portrayal, along these lines space-settled. The space settled portrayal required 3n factors while inward deals with technique required n(n - 1)=2 parts. Cartesian headings delineation is utilized.

The arrangement of directions are spoken to as x = (x1, y1, z1, x2, y2, z2, .......) and V (x) is the potential vitality, then the goal is to discover x that limit the V (x).

**FitnessFunction = MIN (V (x))**  
(3)

Figure 1 describes the flow of process used. The half of the initial population is getting from PDB-Database using PSIBLAST and second half initialize randomly. After initializing the population, fitness is evaluated and then, update or generate the new population according to DIWPSO till the termination criteria is met.

![Fig. 1: Energy minimization methodology using DIWPSO.](image-url)

The initial half population is getting from PSI-BLAST and the half population (dihedral angles (φ, ψ)) is generated randomly, with each sampled uniformly between [-180°, 180°]. The conformation is considered as α-helical when the dihedral angles (φ, ψ) fall in the range [-60 ± 45°, -50 ± 45°] [23]. e.g. a four nucleotides protein is represented by real string of length 12.

<table>
<thead>
<tr>
<th>x1</th>
<th>y1</th>
<th>z1</th>
<th>x2</th>
<th>y2</th>
<th>z2</th>
<th>x3</th>
<th>y3</th>
<th>z3</th>
<th>x4</th>
<th>y4</th>
<th>z4</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.47</td>
<td>3.35</td>
<td>4.16</td>
<td>-2.32</td>
<td>-4.17</td>
<td>-5.64</td>
<td>-1.86</td>
<td>-3.89</td>
<td>-1.23</td>
<td>5.79</td>
<td>2.49</td>
<td>3.66</td>
</tr>
</tbody>
</table>

There are two cases of termination condition are apply in DIWPSO. The first condition based on the convergence on generation that is when The desired generation turnout, the algorithm terminated. The second condition apply when the best fitness value get in the population (referred to as convergence on global best). The algorithm terminates If the ratio between the best value of the Nth previous generation, where N is specified by the user, and the best value of the current generation is larger than a specified value α (0 < α ≤ 1).

### VI. Results Analysis And Discussion
This section discusses the environment setup, parameter setting, structure and minimum energy and cross validation of results. DIWPSO algorithm apply on the molecular system which consists of five amino acids Tyrocin, Veline, Glutemine, Lysine and Arginine shown in Figure 2 for the chemical structure), one sodium ion (Na+), three chlorine ions (CL−), and water as the solvent with 921 water molecules placed in a periodic box with an initial dimension of 3nm on each side. The results are compared with the Genetic Algorithm, Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO) and Classical Differential Evolution (DE) algorithms.

![Fig. 2: Simplified diagram of test molecule consisting of five amino acids: H-VAL-TYR-ARG-LYS-GLN-O.](image)

The motive of the proposed approach is to set the parameter in the string for energy minimization in protein structure prediction. Molecule position in a protein fold encode in string through set the optimize parameter. For calculation of minimum energy we compute its fitness value by AMBER12. We are choosing 38 parameters for energy minimization in this experiment. Also include temperature values, the number of time steps computed in each AMBER12 stage. Maxwell distribution is used for sampling the initial atomic velocities at the particular temperature. In a real value string encode these parameters with length of 114 (3 * 38).

A. Environment setup
For energy computation The AMBER12 [13] package is used, with the parameter setting as imin=1, maxcyc=200, ncycle=75, ntr=10, ntb=0, igb=0, cut=12. Implementation of All the algorithms in C++ in Linux environment and executed on Intel core i5 with 3.2 GHz processor with 4GB of RAM.

B. Required Parameter
Important task is to set the Parameter for algorithms. Taken 50 population size for initial approximation and it can be iterating up to 12000 generation for each algorithms. 3n; is the dimension where n is number nucleotides. Table II summarizes the default parameter setting for each algorithm.

<table>
<thead>
<tr>
<th>SN</th>
<th>Algorithm</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>GA</td>
<td>CR=0.8, MR = 0.2</td>
</tr>
<tr>
<td>2.</td>
<td>DE</td>
<td>CR=0.8, F = 0.6</td>
</tr>
<tr>
<td>3.</td>
<td>PSO</td>
<td>C1, C2 =1.5, w = 0.7</td>
</tr>
<tr>
<td>4.</td>
<td>ABC</td>
<td>NP=30, FOOD SOURCE=NP/2</td>
</tr>
<tr>
<td>5.</td>
<td>DIWPSO</td>
<td>C1, C2 =1.5, D = 0.7</td>
</tr>
</tbody>
</table>

C. Minimum energy
All the algorithm are run for 1400 iterations with initial population size of 50. The Figure 3 shows the final minimized energy for the targeted protein sequence discuss earlier. The minimum energy of $-32.65 \times 10^4$ KJ/mole from DIWPSO and the maximum energy is getting through DE $-30.90 \times 10^4$ KJ/mole.
Energy minimization of protein structure using homology modeling and Particle Swarm...

Fig. 3: Minimum energy getting from different algorithms.

The Table III shows the minimum energy getting through DIWPSO for the public available PDBs. The performance of DIWPSO is better in 50% cases.

**TABLE III:** Minimum Energy Comparison of DIWPSO with publicly available protein.

<table>
<thead>
<tr>
<th>PDB ID</th>
<th>Actual Energy (KJ/mole)</th>
<th>Energy getting through DIWPSO (KJ/mole)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1FTJ</td>
<td>-104132.86</td>
<td>-109258.56</td>
</tr>
<tr>
<td>1TIS</td>
<td>-133269.75</td>
<td>-116491.57</td>
</tr>
<tr>
<td>1PB7</td>
<td>-100296.00</td>
<td>-100352.89</td>
</tr>
<tr>
<td>1S7Y</td>
<td>-115314.22</td>
<td>-116170.86</td>
</tr>
<tr>
<td>1S9T</td>
<td>-136848.94</td>
<td>-124180.05</td>
</tr>
<tr>
<td>2FTJ</td>
<td>-133161.39</td>
<td>-136154.82</td>
</tr>
<tr>
<td>2TIS</td>
<td>-192040.50</td>
<td>-164142.68</td>
</tr>
<tr>
<td>2PB7</td>
<td>-178509.10</td>
<td>-156986.05</td>
</tr>
<tr>
<td>2S7Y</td>
<td>-123616.28</td>
<td>-124646.96</td>
</tr>
<tr>
<td>2S9T</td>
<td>-178026.36</td>
<td>-127477.69</td>
</tr>
</tbody>
</table>

VII. Conclusion

Modified PSO(DIWPSO) presented in This paper. DIWPSO performed with control parameter D. D select Inertia Weight with equal probabilities among four Inertia Weight strategies in each approximation. We have performed the minimization of total empirical energy of protein structure using DIWPSO. We found that DIWPSO outperform in energy minimization of the targeted protein sequence over GA, DE, PSO and ABC algorithms. The result are also compared with public available PDBs and it is found that DIWPSO perform better in 50% cases. The advance Technique of the algorithms is the future scope for big size protein structure; it is also believe that by merging more effective inertia weight strategy may given more better results with higher efficiency.

References


DOI: 10.9790/0661-1906042431 www.iosrjournals.org
Energy minimization of protein structure using homology modeling and Particle Swarm Optimization with Dynamic Inertia Weights.  

References:


