

Write Up

January 1, 2015

1 Problem Statement and Goal

Consider a $n \times n$ square grid $G(V, E)$. Note that the grid has n^2 vertices and $2n(n-1)$ edges. There is a protein chain $P(V_P)$ of length n^2 . Let $V_P = [1, n^2]$ be the vertex set of P . We have a weight function for P . The weight function is, $w : V_P \times V_P \rightarrow R$. In the protein folding model, $w(u, v) = k$ tells us that k amount of energy will be added to the system if vertex u and v ever come in contact (meaning that they are adjacent in the grid but not in the protein chain). (For (u, v) where u and v are adjacent in the chain, the value $w(u, v)$ is unimportant, as we will see later.)

Compact folding(Valid Configuration): A self avoiding walk of length n^2 in G corresponds to a compact folding of a chain of length n^2 onto the grid. Such a walk defines a bijective map $f : V_P \rightarrow V$ describing a position for a chain point on the grid.

Energy of a configuration: For a configuration c , its energy $E(c)$ is defined as

$$E(c) = \sum_{(u,v) \in E} w(f^{-1}(u), f^{-1}(v))$$

Note that the weights of pairs (u, v) where (u, v) is an edge of protein chain only add a constant to $E(c)$.

Problem Statement: Given a Grid $G(V, E)$, protein chain $P(V_P, E_P)$, and weight function w , give a polynomial time (possibly randomised) algorithm to find (or give a good approximation to) the minimum energy valid configuration of the protein chain on the grid. We work with the following cases for the weight function w :

1. Unrestricted: $w : V_P \times V_P \rightarrow \mathbb{R}$. The weights can take any real value.
2. Restricted: $w : V_P \times V_P \rightarrow \{-k, 0, k\}$. All weights are in some finite set.

2 One Approach

We define valid configurations of the protein on this grid using the following boolean formulation: For each edge $e \in E$, let X_e be a boolean variable denoting the selection of e in the grid. Some particular selections of exactly $n^2 - 1$ edges out of $|E|$ edges correspond to valid configurations. Lets call the set of valid configurations $S \subset \{0, 1\}^{2n(n-1)}$. We are interested in $C = \text{ConvexHull}(S)$. We are interested in the following question:

Is there a polynomial time separation oracle for C ?

If we have a polynomial time separation oracle for this polytope and a convex function to be optimised over the polytope (The convex function captures the energy of configurations), then we can solve the combinatorial optimisation problem exactly in polynomial time. Also, consider the following boolean integer program:

The grid is denoted by $G(V, E)$. Let $i \in V$ denote a grid point. Also, let N_i be the set of neighbours for a grid point i . Let the boolean variable X_i^k denote that the k^{th} chain point occupies the grid point i . Here $k = 1, 2, \dots, n^2$. We also have the weight function. For $p, q \in [1, n^2]$, w_{pq} denotes the weight for the chain points p and q . Boolean variables Y_{pq} denotes that the chain points p and q are adjacent in the grid (they may also be adjacent in the chain, but that does not matter). The problem is then:

$$\min \sum_{p \in V_P, q \in V_P} w_{pq} Y_{pq}$$

subject to

$$\sum_{i \in V} X_i^k = 1$$

for $k = 1, 2, \dots, n^2$

$$\sum_k X_i^k = 1$$

$$\forall i \in V$$

$$X_i^k \leq \sum_{u \in N_i} X_u^{k+1}$$

$$\forall k \in [1, n^2 - 1], \forall i \in V$$

$$Y_{pq} \geq X_i^p + X_j^q - 1$$

$$\text{for } (i, j) \in E, \forall p, q \in V_P$$

$$Y_{pq} \leq \sum_{(i,j) \in E} (X_i^p + X_j^q - 1)$$

$$\forall p, q \in V_P$$

In the objective sum, we are also considering the weights of pairs which are adjacent in the chain too and not only in the grid. But this only adds a constant to the objective function.

Let S' be the set of feasible points for this IP. Let $C' = \text{conv}(S')$ be the polytope formed by the solutions as vertices. We are interested in the following question:

Is there a projection of C' that gives C ? If 'yes', then what happens to the objective function under the projection.

Since, we want to optimise a convex function over C , we want to see if the "projected function" can be convexified.

3 Another Approach: Simulation

Rather than restricting to the grid, we can ask for the most stable configuration for the chain in the 2D space. Now, there are infinite configurations for the chain in the 2D space. But if we form a canonical representation, many similar configurations can be considered as one. The energies of the configuration allows us to form a distribution over the configurations. We simulate this distribution using Metropolis Algorithm i.e we construct a Markov Chain that converges to this distribution. This approach has been followed using the random energy model which says that the weight function gives uniformly random weights to each pair of chain points. Once you have the weights sampled randomly, you know start from the straight chain and run the metropolis algorithm. In 2D, 2 kind of local moves have been defined.