

Three-Dimensional Reconstruction Using the Encounter Probabilities

May 28, 2015

1 Using inverted probabilities assuming a Rouse chain

The procedure relies on the nearest neighbor connectivity. As with the previous method, we assume a Rouse polymer encounter probability formula, which drops with the distance between beads on the chain. The main observation here is that if we estimate the nearest neighbors encounter probability for each bead, we can construct the whole connectivity map of the polymer. It will later be left to show that the non-nearest neighbor encounter probability will serve as correction factors.

1.1 Smoothing the encounter signal

To locate prominent features of the signal, we have to reduce noise accompanying the experimental data. There are several ways to smooth the encounter probability data. We list few plausible methods. The smoothing is performed on the encounter frequency signal, before calculating the encounter probability.

1.1.1 Averaging

Due to the structure of the encounter frequency matrix, we have to choose a proper smoothing kernel. To take into account the structure of the chain, smoothing is performed with the following kernel

1.1.2 Iterative Gaussian Smoothing

We start by iteratively applying Gaussian kernel smoothing with increasing variance for several iterations. The number of iteration is determined by the numerical accuracy of the machine used, which produces noise after several iterations, depending on machine precision, which prevents us from applying the filter indefinitely.

One possible numerical criterion for the presence of noise, is the change in the number of local maxima the signal has over all values of variance chosen in the smoothing procedure. We then choose the variance value that give rise to the minimal number of maximal points.

1.1.3 $\mu|\lambda$ smoothing

Taubin's $\mu|\lambda$ smoothing procedure [2] works by sequentially applying Gaussian smoothing with alternate positive and negative scales factors λ and μ respectively, for N iterations. In fact, any type of weighting other than the Gaussian can be applied, with the condition that the weights sum to one, and that no vertex is included in its own neighborhood.

At first we compute v'_i from v_i by adding to it an averaging of its neighborhood with a positive scale factor λ

$$v'_i = v_i + \lambda \sum_{j \in i^*} (v_j - v_i) w_j$$

Then, we compute v''_i from v'_i using a negative scale factor $\mu < -\lambda < 0$

$$v''_i = v'_i + \mu \sum_{j \in i^*} (v'_j - v'_i) w_j$$

with w_i the weighting mask, i^* are the indices of the vertex i neighborhood excluding itself.

In practice, we perform this by two steps convolution over N iterations with a $n_v \times n_v$ square convolution kernel W . The kernel is changed such that $W_{n_v/2, n_v/2} = 0$; $W_{ij} \leftarrow W_{ij} / \sum \sum W_{ij}$. For an input signal V , $V' = V + \lambda(V * W - V)$ and $V'' = V' + \mu(V' * W - V')$, over N iterations.

The variables μ, λ and N are determined by solving the inequalities

$$\begin{aligned} ((\lambda - \mu)^2 / (-4\lambda\mu))^N &< 1 + \delta_{pB} \\ ((1 - \lambda k_{sB})(1 - \mu k_{sB}))^N &< \delta_{sB} \end{aligned}$$

with the constraints:

$$N > 0, \quad 0 < \lambda < -\mu, \quad \lambda < \frac{1}{k_{sB}}, \quad \frac{1}{\lambda} + \frac{1}{\mu} = k_{pB}$$

The filter's parameters are then k_{sB}, k_{pB} . We have to change the shape of the neighborhood for the convolution operations, to take into account the shape of the encounter frequency vs. distance matrix.

1.2 Smoothing by solving the inverse source problem

A fundamental solution to the 1D heat (diffusion) equation is the Gaussian

$$G(x, t, y, \tau) = \frac{H(t - \tau)}{\sqrt{4\pi(t - \tau)}} \exp\left(-\frac{(x - y)^2}{4(t - \tau)}\right)$$

with H the Heaviside step function. It has been shown that in the generalized connectivity Rouse model the distribution of the vectors between beads is a Gaussian. This motivates the use of a combination of the fundamental solution to smooth the HiC data. Given the noisy observations of the HiC we now wish to find a smooth solution to the heat equation that represents the data in the least-square sense.

The HiC encounter signals are non monotonic which hold structural information about the underlying average polymer model. It is therefore not enough to only smooth the curves with the solution of the heat equation but also to allow flexibility of the solution to account for this non monotonicity.

We therefore add a theoretical source term, $r(t)$, in the heat differential equation with which we will smooth the HiC signals. The problem now becomes to find the unknown function $r(t)$ given the noisy measurements $u(x, t)$, which is the inverse source problem.

We start by imagining that each one of the N encounter probability signals represents the temperature of a metal bar, initially hotter than its surrounding. The distance between beads will be thought of as time in the heat equation, and we solve the problem for 3 space points, the middle one being our encounter signal, while the other two will be assigned the boundary conditions.

1.2.1 formulation of the problem

In this subsection I present the formulation given by [1] to solve the inverse time-dependent source problem. let $T > 0$ be a fixed number and let $D_T =$

$\{(x, t) : 0 < x < 1, 0 < t \leq T\}$. consider the problem of finding the temperature $u(x, t)$ and the time-dependent heat source $r(t)$ which satisfy the heat conduction equation

$$u_t = u_{xx} + r(t)f(x, t), \quad (x, t) \in D_T$$

subject to the boundary conditions

$$u(0, t) = u(1, t), \quad u_x(0, t) + \alpha u(0, t) = 0, \quad 0 \leq t \leq T,$$

with $\alpha \neq 0$. The initial condition

$$u(x, 0) = \phi(x), \quad 0 \leq x \leq 1$$

and the additional condition

$$\int_0^1 u(x, t) dx = E(t), \quad 0 \leq t \leq T$$

1.3 The steps

For a HiC read of N equally spaced segments.

1. Obtain the encounter histogram of the HiC data.
2. Arrange the encounters as 3D matrix E , with x -axes as bead index $b = 1, \dots, N$, y -axes as the distance $d = -(N - 1), \dots, (N - 1)$ along the chain, and the z -axes the number of encounters $E(b, d)$. Arranging the HiC data from left to right, we distinguish between 'right' encounter matrix, $e_r(b, d)$ for $d > 0$ and 'left' encounter matrix $e_l(b, -d)$ for $d < 0$ such that $E = e_l(b, -d) \oplus e_r(b, d)$ which is $N \times 2N$ matrix.
3. Smooth $e_r(b, d)$ and $e_l(b, -d)$ using the procedure described in.
4. Normalize each row of the smoothed encounter histogram to get probabilities [Read about bias]

2 Reconstruction

3 Ideas and preliminary work

This section contains preliminary ideas and work. some sections are not edited and the information in them should be double-checked.

3.1 Using optimization methods

Given an encounter probability profile of the first bead in a Rouse chain of N bead, we use the expected encounter model in the Rouse chain to give estimate on the 3D organization of the chain. Given the encounter profile e , we minimize the goal function of the form

$$\min\{f(d)\} \quad (1)$$

$$Ad \leq b \quad (2)$$

$$A_{eq}d = b_{eq} \quad (3)$$

$$l_b \leq d \leq u_b \quad (4)$$

$$c(d) \leq 0 \quad (5)$$

$$c_{eq}(d) = 0 \quad (6)$$

where d is a vector of the distances of bead from bead 1, $f = \sum d_j$, $A \in M_{[N \times N]}$ is a matrix such that

$$A_{ii} = \begin{cases} -1 & \text{if } \exists j \quad ; \|e_j - e_i\| \leq \epsilon \\ 0 & \text{else} \end{cases} \quad (7)$$

$$A_{ij} = \begin{cases} 1/|J| & \text{if } \|e_j - e_i\| \leq \epsilon \\ 0 & \text{else} \end{cases} \quad (8)$$

for small ϵ of choice, and J is the group of all indices j such that $\|e_j - e_i\| \leq \epsilon$ for each i . The vector b is the zero vector, $b = 0$, l_b is the lower bound, set to be a vector of all ones, u_b is the upper bound set to be a vector of all N . The non linear constraint

$$c(d) = \sum_{i=1}^N \left(\frac{d_i^{-1.5}}{\sum_{i=1}^N d_i^{-1.5}} - e_i \right)^2 \quad (9)$$

3.2 using delta functions

For the sake of simplicity, we will look at the encounter probability from the view point of the first bead in a chain of N beads. To take into account the structure of the chain we take the probability density function of the vector between any two beads

$$Pr(R_1 - R_j = r) = \left(\frac{3}{2\pi b^2 |1 - j|} \right)^{3/2} \exp \left(-\frac{3r^2}{2|1 - j|b^2} \right)$$

and for the encounter $R_1 = R_j$ we get

$$Pr(R_1 = R_j) = \left(\frac{3}{2\pi b^2 |1 - j|} \right)^{3/2}$$

Now assume that bead 1 and j are connected by harmonic spring. Then, we make the additional changes to the encounter probability by introducing the random variable δ_{1j} , which indicates whether bead 1 and j are connected. The probability of encounter then becomes

$$Pr(R_1 = R_j) = (1 - \delta_{1j}) \left(\frac{3}{2\pi b^2 |1 - j|} \right)^{3/2} + \delta_{1j} \left(\frac{3}{2\pi b^2} \right)^{3/2}$$

or

$$Pr(R_1 = R_j) = \left(\frac{3}{2\pi b^2} \right)^{3/2} \left(\frac{1 - \delta_{1j} (1 - |1 - j|^{3/2})}{|1 - j|^{3/2}} \right)$$

Given a signal, $e_1(d)$ of HiC encounter probability for bead 1 with beads $d = 2, \dots, N$, we want to estimate the random vector

$$\Delta = [\delta_{13}, \delta_{14}, \dots, \delta_{1N}]$$

For this end, we calculate the maximum likelihood estimator for δ
[Unfinished]

3.3 Using structural distribution function

We first make the assumption that the encounter probability of beads within a loop is well approximated with that of a chain. From the point of view of bead 1. The probability to encounter bead j is given by a superposition of all distances possible between the two beads times the probability of their encounter in the context of a chain. Let the distance between bead 1 and j be δ_{1j} . Then,

$$Pr(R_1 - R_j = 0) = Pr(\delta_{1j} = 1)f(1) + Pr(\delta_{1j} = 2)f(2) + \dots + Pr(\delta_{1j} = j-1)f(j-1)$$

with

$$f(d) = \left(\frac{3}{2\pi b^2 d} \right)^{3/2}$$

which can be written as $Pr(R_1 - R_j = 0) = \langle p, f \rangle > 0$, and indicates that the vectors p and f are independent, as expected. Note that we sum up to the maximal possible distance along the chain between the beads, i.e $j - 1$.

Treating δ as a continuous variable, we set the first bead to be R_0 , and the maximal distance along the chain becomes j . We set $Pr(R_0 - R_j = 0) = P_{0j}$, thus

$$Pr_{0j} = \int_0^j p(\delta) f(\delta) d\delta = \left(\frac{3}{2\pi b^2} \right)^{3/2} \int_0^j \frac{p(\delta)}{\delta^{3/2}} d\delta$$

Using integration by parts we get

$$Pr_{0j} = \left(\frac{3}{2\pi b^2} \right)^{3/2} \left[1 - 2 \frac{p(j)}{\sqrt{j}} + 2 \int_0^j \frac{p^{(1)}(\delta)}{\sqrt{\delta}} d\delta \right] \quad (10)$$

where here we assume that $\lim_{\delta \rightarrow 0} \frac{p(\delta)}{\delta} = 1$, and $p^{(k)}$ signifies the k^{th} derivative of $p(\delta)$. This is a reasonable assumption, since the probability that two adjacent beads are connected as the distance decreases should be 1.

Continuing with the integration by parts we get

$$\left(\frac{3}{2\pi b^2} \right)^{3/2} \left[1 - 2 \frac{p(j)}{\sqrt{j}} + 2[2p^{(1)}(j)\sqrt{j} - 2 \int_0^j \frac{p^{(2)}(\delta)}{\sqrt{\delta}} d\delta] \right]$$

And in the limit we have the formula:

$$Pr_{0j} = \left(\frac{3}{2\pi b^2} \right)^{3/2} \sum_{k=1}^{\infty} (-1)^k p^{(k-1)}(j) \left(\sqrt{j} \right)^{2k-3} \frac{2^k}{\prod_{i=1}^k (2i-1)}$$

We can now state some assumptions regarding the function $p(\delta)$.

1. $p(\delta)$ is infinitely differentiable
2. $p(\delta) > 0 \quad \forall \delta > 0$

References

- [1] A Hazanee, MI Ismailov, D Lesnic, and NB Kerimov. An inverse time-dependent source problem for the heat equation. *Applied Numerical Mathematics*, 69:13–33, 2013.
- [2] Gabriel Taubin. Curve and surface smoothing without shrinkage. In *Computer Vision, 1995. Proceedings., Fifth International Conference on*, pages 852–857. IEEE, 1995.