

III. ПРИКЛАДНЫЕ ЗАДАЧИ РАСПРЕДЕЛЕННЫХ ВЫЧИСЛЕНИЙ

Стационарные состояния в нелинейной модели переноса заряда в ДНК^{*}

А. П. Афанасьев¹, В. Д. Лахно²,
М. А. Посыпкин¹, В. Б. Султанов²

¹ *Центр Грид-технологий и распределенных
вычислений Института системного анализа РАН*

² *Институт математических проблем биологии РАН*

Распределение зарядовой плотности избыточных зарядов в ДНК (электронов и дырок) играет решающую роль в процессах мутагенеза и канцерогенеза. Задача нахождения этой плотности связана с расчетом диагональных элементов матрицы плотности дырок (электронов). Таким образом, проблема сводится к расчету спектра состояний дырок (электронов) в нуклеотидных последовательностях. Рассматриваются подходы к численному решению стационарного дискретного нелинейного уравнения Шредингера, описывающего в рамках квантово-классической модели переноса заряда в ДНК взаимодействие носителя заряда с классическими колебаниями нуклеотидных пар.

1. Модель переноса заряда в ДНК

Носителями заряда в ДНК являются дырки (катион-радикалы). Перемещение дырки по одной цепи ДНК, состоящей из N нуклеотидов, задается вещественной трехдиагональной матрицей

^{*} Работа выполнена при поддержке Совета по грантам Президента Российской Федерации (№ НШ-5511.2008.9).

$$\hat{H} = \begin{pmatrix} a_1 & b_1 & 0 & \cdots & 0 \\ b_1 & a_2 & b_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & b_{N-2} & a_{N-1} & b_{N-1} \\ 0 & \cdots & 0 & b_{N-1} & a_N \end{pmatrix},$$

где a_n — потенциалы окисления нуклеотидов, b_n — матричные элементы перехода дырки между соседними нуклеотидами. В квантово-классической модели переноса заряда вводится взаимодействие дырки с классическими колебаниями нуклеотидов, что в адиабатическом приближении приводит к стационарному дискретному нелинейному уравнению Шредингера (ДНУШ) [1]:

$$\lambda \psi_n = -\kappa_n |\psi_n|^2 \psi_n + a_n \psi_n + b_{n-1} \psi_{n-1} + b_n \psi_{n+1}, \quad (1)$$

где $\vec{\psi} = (\psi_1, \psi_2, \dots, \psi_N)$ — вектор состояния дырки ($\|\vec{\psi}\| = 1$), величина κ_n определяет взаимодействие дырки с колебаниями нуклеотидов на n -м сайте. Здесь мы считаем векторы вещественными $\vec{\psi} \in R^N$.

Имея для данной олигонуклеотидной цепочки полный набор стационарных состояний $\vec{\psi}^{(k)}$, т. е. решений уравнения (1), и их энергий λ_k , $k = 1, \dots, M$, можно при заданной температуре T составить статистическую сумму $Z = \sum_k e^{-\lambda_k/T}$ и по соответствующему ансамблю Гиббса найти заселенности сайтов, т. е. распределение заряда

$$P(n) = \sum_k |\psi_n^{(k)}|^2 e^{-\lambda_k/T} / Z. \quad (2)$$

2. Экстремумы функционала полной энергии

Решениями уравнения (1) являются векторы $\vec{\psi}$, в которых функционал полной энергии

$$F(\vec{\psi}) = \left(\hat{H} \vec{\psi}, \vec{\psi} \right) - \frac{1}{2} \sum_{n=1}^N \kappa_n |\psi_n|^4 \quad (3)$$

имеет экстремум при условии нормировки

$$\|\vec{\psi}\| = 1. \quad (4)$$

Решения, в которых экстремум является минимумом, назовем устойчивыми. Характер экстремума $\vec{\psi}^{(k)}$ определяется по матрице

$$\hat{L} = \left(\frac{\partial^2 F}{\partial \psi_i \partial \psi_j} - 2\lambda_k \delta_{ij} \right)$$

— если проекция \hat{L} на касательную гиперплоскость к сфере в точке экстремума положительно определенная матрица, то это точка минимума, иначе это либо седловая точка, либо максимум. В настоящей работе рассматривались две задачи: поиск глобального минимума функции (3) при ограничении (4) и поиск совокупности стационарных точек.

3. Решение с помощью рекуррентных соотношений

Обобщение на нелинейный случай рекуррентных соотношений для трехдиагональной матрицы \hat{H} позволяет находить вещественные собственные векторы. Определим последовательность полиномов

$$p_n(\lambda, t) \quad (b_N := 1): \quad p_0(\lambda, t) = 0, \quad p_1(\lambda, t) = t, \quad \text{для } n = 1, \dots, N:$$

$$p_{n+1}(\lambda, t) = \left((\lambda - a_n + \kappa_n |p_n(\lambda, t)|^2) p_n(\lambda, t) - b_{n-1} p_{n-1}(\lambda, t) \right) / b_n.$$

Векторы $\vec{\psi}(\lambda, t) = (p_1(\lambda, t), p_2(\lambda, t), \dots, p_N(\lambda, t))$ будут нормированными решениями уравнений (1) если

$$p_{N+1}(\lambda, t) = 0, \quad \|\vec{\psi}(\lambda, t)\|^2 = \sum_{n=1}^N |p_n(\lambda, t)|^2 = 1. \quad (5)$$

Параметр t можно считать: $0 < t < 1$, интервал для λ : $\lambda_{\min} \leq \lambda \leq \lambda_{\max}$ можно оценить для данной задачи. Сведение задачи к двум параметрам (λ, t) позволяет найти интервалы, содержащие решения и в них численно решать систему (5). Так были найдены все стационарные решения для цепочек $GG\dots G$ при $N \leq 11$. Из расчетов следует описание всех решений для $GG\dots G$ при $\kappa = 4$: $N - 2$ минимума, $N - 3$ седловые точки индекса 1 и остальные $N - 1$ решений — это деформации от $\kappa = 0$ до $\kappa = 4$ $N - 1$ линейных решений — всего $M = 3N - 6$ решений ($N \geq 5$). На основе найденных значений по формуле (2) было вычислено распределение заряда (см. [2]).

4. Решение с использованием метода проекции градиента

Применяемый алгоритм формирует список локальных стационарных точек Y . Первоначально список пуст. Алгоритм работает по следующей схеме.

1. Выбрать точку $x \in S_N$.
2. Применить алгоритм L локальной оптимизации с точкой x в качестве начальной.
3. Добавить результат оптимизации, точку $y = L(x)$ в список Y локальных минимумов.
4. Перейти к шагу 1, если не превышен лимит количества итераций или времени.

Так как численный алгоритм находит точку минимума приближенно, то необходимо ввести параметр δ для проверки того, что найденные точки соответствуют одному минимуму. Точка y добавляется в список Y , только если она не находится в δ -окрестности ранее найденных решений. Просматривается весь список Y и, если в нем находится точка z , такая что $\|y - z\| \leq \delta$, то сравниваются значения целевой функции в этих точках.

Если $f(z) \leq f(y)$, то точка y не добавляется в список, в противном случае она замещает в списке точку z . Если для всех точек $z \in Y$ справедливо $\|y - z\| \geq \delta$, то точка y добавляется к списку.

При больших значениях параметров δ возможно отождествление различных минимумов, а при малых — наоборот, могут появиться ложные дополнительные минимумы. В рассматриваемом случае выбиралось достаточно большое значение δ , которое уменьшалось до тех пор, пока количество стационарных точек не переставало увеличиваться.

Применяемые алгоритмы локальной оптимизации останавливаются, если норма градиента меньше заданной точности ε . Найденная точка может при этом оказаться минимумом, седловой точкой и, в редких случаях, — точкой максимума. Можно проводить дополнительный анализ для того, чтобы определить характер найденных точек.

Если не делать каких-либо предположений о расположении экстремумов, то целесообразно проводить генерацию равномерно по гиперсфере S_N . Известно, что если точки x_1, \dots, x_n — случайные числа, распределенные по нормальному закону с математическим ожиданием 0 и дисперсией 1, то точки $x/\|x\|$, $x = (x_1, \dots, x_n)$ распределены равномерно по поверхности гиперсферы с единичным радиусом. Также возможно применение метода *генерации по осям*, когда в качестве начальных используются точки пересечения

осей координат с единичной сферой, т. е. векторы с одним ненулевым компонентом, равным ± 1 . В качестве локального алгоритма применялся метод наискорейшего спуска по направлению проекции анти-градиента[3].

Коэффициенты функции (1) формировались в соответствии с [1]. Эксперименты проводились для различных значений параметра n и способа генерации последовательности ДНК.

4.1. Однородные последовательности

Рассматриваются однородные последовательности Гуанинов $GGG...G$. В таблицах 1–4 приведены следующие величины:

min — минимальное значение функции f в стационарной точке;

max — максимальное значение функции f в стационарной точке;

N_s — общее число полученных стационарных точек.

В табл. 1 приведены данные, полученные при генерации по осям. В табл. 2 — при генерации совокупности 50000 равномерно распределенных точек по поверхности гиперболы S_n .

Глобальный минимум надежно находится обоими алгоритмами. Метод генерации по осям находит больше стационарных точек, так как точки пересечения осей с гиперсферой находятся в окрестности глобальных минимумов. С другой стороны, метод случайного поиска находит дополнительные стационарные точки, которые не находятся при генерации по осям. Поэтому наибольшее число стационарных точек было получено при совместном применении обоих методов генерации начальных точек (табл. 3).

Таблица 1

	50	500	1000
min	-2,868996	-2,868996	-2,868996
max	-2,815513	-2,815513	-2,815513
N_s	96	996	1996

Таблица 2

	50	500	1000
min	-2,868996	-2,868996	-2,868996
max	-2,815513	-2,834931	-2,834931
N_s	92	972	1979

Таблица 3

	50	500	1000
min	-2,868996	-2,868996	-2,868996
max	-2,815513	-2,815513	-2,815513
N_s	97	999	2001

Можно также заметить, что значения целевой функции не зависят от размерности в пределах шести знаков после запятой. Это является следствием того, что в найденных стационарных точках, как правило, небольшой участок вектора заметно больше нуля, в то время как остальные компоненты очень близки к нулю и практически не оказывают влияния на значение целевой функции.

4.2. Случайные последовательности

На случайных последовательностях также наилучшие результаты дает сочетание генерации по осям и равномерно по сфере. Результаты для случайных цепочек различной длины приведены в табл. 4.

Таблица 4

	50	500	1000
min	-3,634694	-3,953951	-4,219602
max	-1,66716	-2,545405	-2,590549
N_s	25	139	82

Заключение

В работе рассмотрена задача поиска экстремума функционала полной энергии в нелинейной модели переноса заряда в молекуле ДНК. Предложены точные, основанные на рекуррентных соотношениях, и стохастические подходы к определению минимума функционала. Для стохастических подходов эксперименты показали, что увеличение числа генерируемых начальных приближений повышает количество найденных стационарных точек. При этом для последовательностей длиной 1000 и более целесообразно генерировать 10^6 точек и более. Поэтому в ближайшем будущем планируется распараллелить предложенный алгоритм.

Литература

1. Лахно В. Д., Фиалко Н. С. Перенос заряда в ДНК на большое расстояние // Компьютеры и суперкомпьютеры в биологии. М.: ИКИ, 2002.
2. Лахно В. Д., Султанов В. Б. Распределение заряда в ДНК при конечной температуре // Математическая биология и биоинформатика: II Международная конференция, Пушино, 7–13 сентября 2008 г.: Доклады / Под ред. В. Д. Лахно. С. 17–19.
3. Поляк Б. Т. Введение в оптимизацию. М.: Наука, 1983.