

**АКАДЕМИЯ НАУК СССР
НАУЧНЫЙ ЦЕНТР БИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЙ
НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ВЫЧИСЛИТЕЛЬНЫЙ ЦЕНТР**

ПРЕПРИНТ

В.Ю.ЛУНИН, А.Г.УРЖУМЦЕВ

**КОНСТРУИРОВАНИЕ ПРОГРАММ
ДЛЯ УТОЧНЕНИЯ АТОМНОЙ СТРУКТУРЫ
МАКРОМОЛЕКУЛ НА БАЗЕ АЛГОРИТМОВ
БЫСТРОГО ПРЕОБРАЗОВАНИЯ ФУРЬЕ
И БЫСТРОГО ДИФФЕРЕНЦИРОВАНИЯ**

ПУЩИНО • 1983

УДК 517.52:681.3.06:548.73

В работе сформулирован общий подход к уточнению атомной структуры макромолекулы. Построен алгоритм, позволяющий осуществлять цикл уточнения за время, линейно зависящее от размеров молекулы.

The structure refinement can be considered as a minimization problem for some function $f(x)$ which depends on large number of parameters. A new type of such a function is proposed incorporating phase probability distributions.

∇f and $\nabla^2 f \cdot \vec{e}$ calculations are required for a step of $f(x)$ minimization by conjugate gradients techniques. (Here ∇f , $\nabla^2 f$ and \vec{e} are gradient, matrix of second derivatives and any direction in parameter space). Kim et al. have proposed an algorithm in which the cost of calculation of all ∇f and $\nabla^2 f \cdot \vec{e}$ components is nearly equivalent to the cost of single $f(x)$ value computation. It is to be emphasized that in this algorithm $\nabla^2 f$ is full matrix of second derivatives without any approximations. The algorithm is adapted for protein refinement problem.

ВВЕДЕНИЕ

Данная работа посвящена выработке общего подхода к разработке программ для уточнения структур больших биомолекул. Говоря об уточнении структуры, мы подразумеваем, что нам уже известна некоторая модель изучаемого объекта и речь идет о корректировке значений параметров этой модели таким образом, чтобы добиться удовлетворительного соответствия данным рентгеновского рассеяния и некоторым дополнительным требованиям. В настоящее время существует ряд программ для уточнения белковых структур, различающихся способами описания модели объекта, критериями соответствия модели экспериментальным данным и способами учета дополнительной (стереохимической) информации. В § 1 мы формулируем задачу уточнения атомной модели структуры в общем виде. При этом в известной мере стирается граница между понятиями уточнения в прямом и обратном пространствах, а также объединяются жесткий и нежесткий способы введения стереохимических ограничений.

Существенной особенностью уточнения структур макромолекул является большое число параметров, подлежащих уточнению. Это приводит к серьезным вычислительным сложностям, и приемлемыми для практической реализации являются, как правило, алгоритмы уточнения, объем вычислений в которых линейно растет с ростом размеров уточняемого объекта. Особенно существенным использование таких алгоритмов становится по мере увеличения размеров исследуемых белков. Создание алгоритма Кули-Тьюки быстрого преобразования Фурье и разработка Кимом и др. алгоритма быстрого дифференцирования позволили в самом общем случае построить алгоритм уточнения модели, затраты на один цикл которого растут практически линейно с ростом размеров молекулы. Некоторые общие вопросы минимизации функций большого числа переменных рассмотрены в §2, в котором центральное место – алгоритм Кима, Нестерова и Черкасского, позволяющий вычислять все n компонент градиента произвольной функции $f(x_1, \dots, x_n)$ "за ту же цену", что и однократное вычисление функции. В § 3 мы показываем, как практически построить такой алгоритм в задаче уточнения атомной модели структуры. Для некоторого частного

случая критерия уточнения такого рода алгоритм (позволяющий рассчитывать градиент за время, пропорциональное времени вычисления функции) был предложен ранее Агарвалом /1/. Однако "взгляд сверху" на проблему построения такого алгоритма делает его построение более прозрачным и позволяет построить более эффективный алгоритм.

§ 1. Постановка задачи уточнения модели

1.1. Модель исследуемого объекта

Экспериментальной основой рентгеноструктурного исследования вещества является рассеяние рентгеновских лучей на электронах атомов. Поэтому в рамках этой методики моделью изучаемого объекта является функция распределения электронной плотности в нем. В данной работе мы ограничимся рассмотрением "атомных" моделей исследуемого объекта, то есть случаев, когда функция распределения электронной плотности может быть представлена в виде суммы электронных плотностей отдельных атомов:

$$\rho(\vec{r}) = \sum_j \rho_j(\vec{r}).$$

При этом предполагается, что распределение $\rho_j(\vec{r})$ электронной плотности в отдельном атоме не зависит от его окружения, то есть не изменяется при образовании химической связи.

Существуют разные способы задания распределения $\rho_j(\vec{r})$. Мы будем предполагать, что каждому атому соответствует вектор параметров $\vec{q}_j = \{q_{j\alpha}\}_\alpha$ определяющий вклад $\rho_j(\vec{r}, \vec{q}_j)$ этого атома в функцию $\rho(\vec{r})$.

Примером введения таких параметров является случай, когда распределение электронной плотности отдельного атома определяется четырьмя параметрами - параметром изотропных тепловых колебаний B_j и координатами центра атома $\vec{z}_j = (x_j, y_j, z_j)$. При этом мы будем предполагать, что нам известно для каждого атома распределение электронной плотности $\rho_j^0(r)$, отвечающее этому атому, покоящемуся в начале координат. Таким образом, моделируя влияние тепловых колебаний путем свертки функции $\rho_j^0(r)$ с гауссовой функцией $(4\pi/B_j)^{3/2} \exp\{-4\pi^2 r^2/B_j\}$, мы имеем

$$\rho_j(\vec{r}) = \int_{R^3} \rho_j^0(|\vec{u} - \vec{z}_j|) (4\pi/B_j)^{3/2} \exp\{-4\pi^2 |\vec{u} - \vec{z}_j|^2/B_j\} dV_{\vec{u}}.$$

При распространенной аппроксимации $\rho_j^0(r)$ одной или несколькими гауссовыми функциями функция $\rho_j(\vec{r})$ также будет иметь вид суммы гауссовых функций.

Мы будем обозначать $\vec{q} = \{\vec{q}_j\}_j$ - совокупность параметров всех атомов, дающих вклад в функцию $\rho(\vec{r}) = \rho(\vec{r}, \vec{q})$, и на-

зывать их атомными параметрами модели. Чтобы определить функцию $\rho(\vec{z}, \vec{q})$, нам, конечно, не нужно явно задавать параметры всех атомов. Достаточно задать параметры атомов асимметричной части структуры, так как остальные могут быть получены из них операциями симметрии. Более того, исходное описание модели может вообще не содержать параметров \vec{q}_j , а может, например, содержать углы внутреннего вращения φ_j и ψ_j аминокислотных остатков со стандартной стереохимией /3/. Для нас будет важно лишь, чтобы была дана информация, по которой мы в состоянии однозначно восстановить параметры $\{\vec{q}_j\}_j$ всех атомов.

1.2. Уточнение модели по данным рентгеновского рассеяния

Под уточнением модели по данным рентгеновского рассеяния понимается обычно изменение параметров модели таким образом, чтобы структурные факторы

$$f_{\vec{s}}^R(\vec{q}) + i f_{\vec{s}}^I(\vec{q}) = \int_V \rho(\vec{z}, \vec{q}) e^{2\pi i(\vec{s}, \vec{z})} dV_{\vec{z}},$$

рассчитанные по модели $\rho(\vec{z}, \vec{q})$, наилучшим образом воспроизвели результат экспериментального определения этих величин. Для того, чтобы выразить численно понятие "наилучшим образом", для каждого узла \vec{s} из некоторого множества узлов \vec{S} решетки обратного пространства вводится критерий $\alpha(f_{\vec{s}}^R, f_{\vec{s}}^I; \vec{s})$ соответствия величин $f_{\vec{s}}^R$ и $f_{\vec{s}}^I$ результату эксперимента. Уточнение модели происходит далее путем минимизации функционала

$$R_x(\vec{q}) = \sum_{\vec{s} \in \vec{S}} \alpha(f_{\vec{s}}^R(\vec{q}), f_{\vec{s}}^I(\vec{q}); \vec{s}) \Rightarrow \min. \quad (1)$$

Мы будем работать с вещественной и мнимой частями (комплексных) структурных факторов $\vec{F}(\vec{s}) = f_{\vec{s}}^R + i f_{\vec{s}}^I$ и считать, что критерии заданы как функции от f^R и f^I . Более обычным в кристаллографии является представление структурного фактора в виде

$$\vec{F}(\vec{s}) = F_{\vec{s}} e^{i\varphi_{\vec{s}}}$$

и задание критериев соответствия как функций, зависящих от $F_{\vec{s}}$ и $\varphi_{\vec{s}}$. Так, в настоящее время обычно говорят об уточнении в обратном пространстве, когда критерий $\alpha(F, \varphi; \vec{s})$ имеет вид /1/:

$$\alpha(F, \varphi; \vec{s}) = \omega(\vec{s})(F - F_0(\vec{s}))^2 \quad (2)$$

либо /2/:

$$\alpha(F, \varphi; \vec{s}) = \omega(\vec{s})(F^2 - F_0^2(\vec{s}))^2, \quad (3)$$

и говорят об уточнении в прямом пространстве, когда критерий имеет вид /3/:

$$\alpha(F, \varphi; \vec{S}) = |F e^{i\varphi} - F_0(\vec{S}) e^{i\varphi_0(\vec{S})}|^2 \quad (4)$$

Здесь $\omega(\vec{S})$ – заданная весовая функция, $F_0(\vec{S})$ – экспериментально определенные модули и $\varphi_0(\vec{S})$ – определенные каким-либо образом фазы структурных факторов. Мы будем тем не менее предполагать далее, при выводе формул в § 3, что критерии переформулированы в терминах f^R и f^I , так как это делает более наглядными получаемые формулы, а сам переход от (F, φ) к (f^R, f^I) не представляет труда.

Критерий (4) обладает тем преимуществом перед критериями (2) и (3), что в нем используется дополнительная информация о фазах $\varphi_0(\vec{S})$. В то же время, если эти фазы $\varphi_0(\vec{S})$ определены ненадежно, то требование "подгонки" фаз $\varphi(\vec{S})$ к фазам $\varphi_0(\vec{S})$ в критерии (4) становится препятствием для правильного изменения модели. Достоинства критериев (3) и (4) можно в какой-то мере объединить, если проводить "подгонку" фаз $\varphi(\vec{S})$ к фазам $\varphi_0(\vec{S})$ с учетом надежности этих фаз. В качестве такого критерия можно предложить критерий:

$$\alpha(F, \varphi; \vec{S}) = \frac{1}{\sigma_I^2(\vec{S})} (F^2 - F_0^2(\vec{S}))^2 - [A(\vec{S}) \cos \varphi + B(\vec{S}) \sin \varphi + C(\vec{S}) \cos 2\varphi + D(\vec{S}) \sin 2\varphi] \quad (5)$$

Здесь

$$P(\varphi, \vec{S}) \sim \exp\{A(\vec{S}) \cos \varphi + B(\vec{S}) \sin \varphi + C(\vec{S}) \cos 2\varphi + D(\vec{S}) \sin 2\varphi\}, \quad (6)$$

полученное каким-либо образом распределение вероятностей для фазы, записанное в унифицированном виде [5], $\sigma_I^2(\vec{S})$ – дисперсия измерения интенсивности $I = F_0^2(\vec{S})$. Если считать, что ошибки в измерении интенсивностей $I(\vec{S})$ распределены нормально и взаимно независимы, то использование критерия (5) может быть проинтерпретировано как определение параметров модели из максимума функции правдоподобия. Отметим, что если, например, распределение (6) определяется по данным изоморфного замещения, то коэффициенты A, B, C, D определяются через модули структурных факторов тяжелоатомных производных и параметры тяжелых атомов, то есть, используя критерий (5), мы включаем в уточнение дополнительно результаты эксперимента с тяжелоатомными производными. Отметим также, что критерий (5) позволяет включить в уточнение часть структурных факторов с известными фазами и часть с неизвестными (для них $A=B=C=D=0$), чего не допускает критерий (4).

Поскольку $\vec{F}(\vec{S}) = \vec{F}(\vec{S})^*$, то относительно критериев $\alpha(f^R, f^I; \vec{S})$ и множества S мы примем естественные предположения, что

$$-S = S \quad (7)$$

(то есть S симметрично относительно начала координат) и

$$a(u, -v; -\vec{s}) = a(u, v; \vec{s}).$$

Отсюда, в частности, следует, что

$$a_1(u, v; \vec{s}) = a_1(u, v; \vec{s}), \quad a_2(u, -v; -\vec{s}) = -a_2(u, v; \vec{s}), \quad (8)$$

где обозначено

$$a_1(u, v; \vec{s}) = \frac{\partial}{\partial u} a(u, v; \vec{s}), \quad a_2(u, v; \vec{s}) = \frac{\partial}{\partial v} a(u, v; \vec{s}).$$

1.3. Ограничения на параметры атомов

При работе с атомной моделью на значения параметров модели обычно накладываются некоторые ограничения, отражающие дополнительную информацию об исследуемом объекте. Можно выделить два подхода к введению таких ограничений.

При жестком введении ограничений дополнительная информация поставляется в виде системы равенств

$$\Psi_\alpha(\vec{q}_1, \vec{q}_2, \dots) = 0, \quad \alpha = 1, 2, \dots, \quad (9)$$

которым должны удовлетворять параметры модели. Примеры таких равенств – требование сохранения стандартных длин связей, величин валентных углов, требование нахождения нескольких атомов в одной плоскости. Другим примером наличия ограничений типа (9) является присутствие в структуре локальной (некристаллографической) симметрии. В этом случае параметры некоторых пар атомов (при этом оба атома могут находиться в асимметричной части ячейки) связаны соотношениями типа

$$B_j - B_k = 0 \\ \vec{z}_j - (G\vec{z}_k + \vec{t}) = 0,$$

где (G, \vec{t}) – локальное преобразование симметрии.

При наличии жестких ограничений атомные параметры \vec{q} не являются независимыми, и в процессе уточнения модели допустимы лишь такие изменения параметров, при которых сохраняются равенства (9). Мы будем считать, что в такой ситуации у нас для описания модели введены обобщенные параметры $\vec{x} = (\vec{x}_1, \dots, \vec{x}_N)$ такие, что:

- а) все параметры \vec{q}_j однозначно определяются по параметрам \vec{x} , то есть известны зависимости $\vec{q}_j = \vec{q}_j(\vec{x})$;
 б) при всех \vec{x} равенства (9) выполняются, то есть

$$\Psi_\alpha(\vec{q}_1(\vec{x}), \vec{q}_2(\vec{x}), \dots) = 0 \quad \alpha = 1, 2, \dots$$

В этом случае, произвольно варьируя параметры \vec{X} , мы все время будем иметь модель, атомные параметры которой удовлетворяют соотношениям (9), и задачу минимизации функционала (1) при условиях (9) можно переформулировать как задачу безусловной минимизации функционала $R(\vec{X}) = R_x(\vec{q}(\vec{X}))$.

Примером использования обобщенных параметров является описание модели белка углами φ_i и ψ_i внутреннего вращения аминокислотных остатков. (О более общем подходе к введению независимых параметров см. /7/).

Другой путь - нежесткое введение ограничений - заключается в том, что дополнительная информация формулируется как требование минимальности некоторого функционала (штрафной функции):

$$R_S(\vec{q}_1, \vec{q}_2, \dots) \Rightarrow \min. \quad (10)$$

Такая постановка задачи возникает, если мы не требуем точного сохранения длин связей, углов и т.п., а хотим лишь, чтобы они как можно меньше отличались от стандартных. Другой пример возникновения условий типа (10) - ряд работ (например, /6/), в которых функционал $R_S(\vec{q})$ имеет смысл энергии конформации молекулы, и уточнение модели проводится так, чтобы по возможности минимизировать энергию.

При нежестком включении ограничений задача уточнения модели формулируется как задача минимизации составного функционала

$$\alpha R_X(\vec{q}) + \beta R_S(\vec{q}) \Rightarrow \min, \quad (11)$$

где α и β - некоторые веса, правильный выбор которых представляет отдельную проблему. После того, как значения параметров α и β выбраны, мы можем убрать их "внутрь" функционалов R_X и R_S .

Заметим, что при уточнении модели могут одновременно использоваться как жесткие, так и нежесткие ограничения, то есть мы можем минимизировать функционал (11) при дополнительных условиях типа (9).

Отметим еще, что помимо параметров модели функционалы $R_X(\vec{q})$ и $R_S(\vec{q})$ могут содержать еще некоторые параметры, уточняемые в процессе минимизации функционала (11). Такими параметрами могут быть коэффициент приведения $F_0(S)$ к абсолютной шкале, параметры "dummy"-атомов (/4/), вводимые для упрощения вида функционала R_S , и т.п.

1.4. Резюме

Итак, мы будем считать далее, что

а) состояние модели описывается набором обобщенных независимых параметров $\vec{X}_1, \dots, \vec{X}_n$;

б) заданы формулы перехода к атомным параметрам модели

$$\vec{q}_j = \vec{q}_j(\vec{x});$$

в) уточнение модели заключается в изменении параметров $\vec{x}_1, \dots, \vec{x}_n$ таким образом, чтобы минимизировать величину

$$R(\vec{x}) = R_x(\vec{q}(\vec{x})) + R_s(\vec{q}(\vec{x})), \quad (12)$$

при этом функционал $R_x(\vec{q})$ имеет вид (1).

§ 2. Минимизация функций большого числа переменных

2.1. Метод сопряженных градиентов

Мы рассмотрим задачу о нахождении точки минимума $\vec{x}^* = \text{arg min } f(\vec{x})$ функции $f(\vec{x})$ большого числа переменных $\vec{x} = (x_1, \dots, x_n)$. Если функция $f(\vec{x})$ квадратичная, то есть

$$f(\vec{x}) = \frac{1}{2} (\vec{x}, H\vec{x}) + (\vec{x}, \vec{b}) + c, \quad (13)$$

(где $(\vec{x}, \vec{y}) = \sum_{j=1}^n x_j y_j$), то формальный ответ дается равенством

$$\vec{x}^* = \text{arg min}_{\vec{x}} f(\vec{x}) = -H^{-1}\vec{b} = \vec{x}^0 - (\nabla^2 f|_{\vec{x}^0})^{-1}(\nabla f|_{\vec{x}^0}), \quad (14)$$

где \vec{x}^0 - произвольная точка, $\nabla f|_{\vec{x}^0}$ - градиент в точке \vec{x}^0 функции $f(\vec{x})$, $\nabla^2 f|_{\vec{x}^0} = H$ - матрица вторых производных (матрица Гесса), вычисленных в точке \vec{x}^0 . Для неквадратичной функции $f(\vec{x})$ мы можем построить в окрестности произвольной точки \vec{x}^k квадратичную аппроксимацию:

$$\begin{aligned} f(\vec{x}) &\approx f(\vec{x}^k) + (\nabla f|_{\vec{x}^k}, \vec{x} - \vec{x}^k) + \frac{1}{2} (\nabla^2 f|_{\vec{x}^k} (\vec{x} - \vec{x}^k), \vec{x} - \vec{x}^k) + \dots = \\ &= \frac{1}{2} (\vec{x}, H_k \vec{x}) + (\vec{x}, \vec{b}_k) + c_k + \dots, \end{aligned}$$

где $H_k = \nabla^2 f|_{\vec{x}^k}$ - матрица Гесса в точке \vec{x}^k . Обобщением формулы (14) служит процедура построения приближений $\vec{x}^0, \vec{x}^1, \dots$ к точке минимума \vec{x}^* по формуле

$$\vec{x}^{k+1} = \vec{x}^k - H_k^{-1} \nabla f|_{\vec{x}^k},$$

представляющая собой не что иное, как решение системы нормальных уравнений

$$\nabla f|_{\vec{x}} = 0$$

методом Ньютона. При надлежащих предположениях относительно

$f(\vec{x})$ доказана сходимость этого метода, и он или его модификации широко используются при уточнении структур низкомолекулярных соединений. Однако при большой размерности n вектора $\vec{x} = (x_1, \dots, x_n)$ попытка практической реализации этой процедуры приводит к значительным вычислительным сложностям уже потому, что требует вычисления n^2 элементов матрицы H_k . Это заставляет использовать вместо матрицы H_k какие-то ее аппроксимации, что ухудшает сходимость.

В последнее время из методов минимизации функций большого числа переменных хорошо зарекомендовал себя в вычислительной практике метод сопряженных градиентов, одна из реализаций которого изложена ниже. Рассмотрим опять квадратичную функцию (13). Пусть найден такой набор векторов $\{\vec{e}_1, \dots, \vec{e}_n\}$, что

$$(\vec{e}_j, H\vec{e}_k) = 0, \quad j \neq k. \quad (15)$$

В этом случае если $\vec{x} = \sum_{j=1}^n \alpha_j \vec{e}_j$, то

$$f(\vec{x}) = \sum_{j=1}^n f(\alpha_j \vec{e}_j),$$

то есть задача минимизации $f(\vec{x})$ распалась на n независимых задач одномерной минимизации:

$$\alpha_j^* = \operatorname{argmin} f(\alpha \vec{e}_j), \quad j = 1, \dots, n;$$

$$\vec{x}^* = \sum_{j=1}^n \alpha_j^* \vec{e}_j.$$

Оказывается, что базис $\{\vec{e}_1, \dots, \vec{e}_n\}$, обладающий свойством (15), может быть построен при помощи следующей процедуры:

- а) пусть \vec{x}^0 - произвольная точка, возьмем $\vec{e}_1 = -\nabla f|_{\vec{x}^0}$;
- б) пусть найдены $\vec{x}^0, \dots, \vec{x}^{k-1}$ и $\vec{e}_1, \dots, \vec{e}_k$, тогда $\vec{x}^k = \vec{x}^{k-1} + \alpha_k \vec{e}_k$, где $\alpha_k = \operatorname{argmin} f(\vec{x}^{k-1} + \alpha \vec{e}_k)$ - решение одномерной задачи минимизации, и $\vec{e}_{k+1} = -\nabla f|_{\vec{x}^k} + \beta_k \vec{e}_k$, где β_k выбрано так, чтобы

$$(\vec{e}_{k+1}, H\vec{e}_k) = 0.$$

Поскольку в процессе построения векторов \vec{e}_k проводится минимизация $f(\vec{x})$ по предыдущим направлениям, то \vec{x}^n является точкой минимума \vec{x}^* квадратичной функции $f(\vec{x})$.

Одним из обобщений указанной выше процедуры на случай неквадратичной функции $f(\vec{x})$ является ее использование с заменой последнего условия, определяющего параметр β , на

$$(\vec{e}_{k+1}, H_k \vec{e}_k) = 0,$$

где $H_k = \nabla^2 f / \vec{x}^k$, то есть
$$\beta = \frac{(\nabla f / \vec{x}^k, H_k \vec{e}_k)}{(\vec{e}_k, H_k \vec{e}_k)}$$

Отметим, что при осуществлении этой процедуры построения точки нам не нужно знать матрицу H_k — нам достаточно уметь вычислять для произвольного вектора \vec{e}_k другой вектор $H_k \vec{e}_k$ (то есть n чисел, а не n^2 , как в матрице H_k). Как будет показано ниже, эта операция может быть проведена существенно меньшими затратами труда, нежели вычисление всех элементов матрицы H_k . Отметим еще, что задача нахождения минимума функции на заданном направлении существенно легче задачи многомерной минимизации, и мы не будем обсуждать ее в этой работе.

Хотя для квадратичной функции метод сопряженных градиентов и позволяет достичь точки минимума за конечное число n шагов, при большой размерности пространства мы не в состоянии сделать все эти шаги. Поэтому целесообразно попытаться ускорить сходимость, проводя минимизацию в более правильной для нашей задачи системе координат (например, выбирая для центров атомов абсолютные координаты вместо относительных). Мы будем предполагать, что в точке \vec{x}^k переход к новым координатам дается равенством

$$\vec{x}' = U_k \vec{x}, \quad (16)$$

где U_k — выбранные нами матрицы, на выборе которых мы не будем здесь останавливаться. Тогда, вводя в процедуру нахождения точки минимума \vec{x}^* преобразования координат (16), приходим к следующему алгоритму построения приближений $\vec{x}^0, \vec{x}^1, \dots$:

а) пусть \vec{x}^0 — произвольно, $\vec{e}_1 = -u_0^2 (\nabla f / \vec{x}^0)$;

б) пусть построены $\vec{x}^0, \dots, \vec{x}^{k-1}, \vec{e}_1, \dots, \vec{e}_k$, тогда

$$\vec{x}^k = \vec{x}^{k-1} + \alpha_k \vec{e}_k, \text{ где } \alpha_k = \arg \min f(\vec{x}^{k-1} + \alpha \vec{e}_k), \quad (17)$$

$$\vec{e}_{k+1} = -u_k^2 (\nabla f / \vec{x}^k) + \beta_k \vec{e}_k, \text{ где } \beta_k = (u_k^2 (\nabla f / \vec{x}^k), H_k \vec{e}_k) / (\vec{e}_k, H_k \vec{e}_k).$$

2.2. Алгоритм быстрого дифференцирования

При практической реализации алгоритма (17) основную вычислительную сложность, на первый взгляд, представляет расчет n компонент градиента $\nabla f / \vec{x}^k$ и n компонент направления $H_k \vec{e}_k$ (матрица H_k — это n^2 вторых производных функции $f(\vec{x})$). На первый взгляд, при большом n эта задача существенно более трудоемка, нежели расчет одного значения функции $f(\vec{x})$. Однако в ряде практических случаев удавалось построить алгоритм (/1/), позволяющий вычислить все n компонент градиента за время, сравнимое с временем вычисления значения функции $f(\vec{x})$. Замечательным является то обстоятельство, что, как показали

Ким и др., для любой функции $f(x)$ можно построить алгоритм вычисления всех компонент градиента $\nabla f|_{\vec{x}}$ за почти ту же цену, что и вычисление функции $f(\vec{x})$. Более того, вычисление вектора $\nabla^2 f|_{\vec{x}} \vec{e}$ также осуществимо за время, сравнимое с временем вычисления $f(\vec{x})$ и независимое от n . Ниже мы приведем построение алгоритма быстрого дифференцирования в виде, удобном для дальнейшей работы.

Рассмотрим ситуацию, когда функция $f(\vec{x})$ вычисляется как суперпозиция функций $F(\vec{y}) = F(y_1, \dots, y_m)$ и $\vec{y}(\vec{x}) = (y_1(\vec{x}), \dots, y_m(\vec{x}))$, то есть когда $f(\vec{x}) = F(\vec{y}(\vec{x}))$. В этом случае

$$\frac{\partial f}{\partial x_j} = \sum_{k=1}^m \frac{\partial F}{\partial y_k} \cdot \frac{\partial y_k}{\partial x_j}, \quad (18)$$

то есть, обозначая

$$\frac{\partial \vec{y}}{\partial \vec{x}} = \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \dots & \frac{\partial y_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial y_m}{\partial x_1} & \dots & \frac{\partial y_m}{\partial x_n} \end{pmatrix}$$

матрицу Якоби, соответствующую замене переменных $\vec{x} \rightarrow \vec{y}(\vec{x})$, и обозначая $\nabla_x f$ градиент функции $f(\vec{x})$ по переменным \vec{x} , можем записать

$$\nabla_x f = \left(\frac{\partial \vec{y}}{\partial \vec{x}} \right)^T \cdot \nabla_y F. \quad (19)$$

Если теперь функция $f(\vec{x})$ вычисляется как более длинная цепочка суперпозиций

$$f(\vec{x}) = F(\vec{y}^N(\vec{y}^{N-1}(\dots \vec{y}^2(\vec{y}^1(\vec{x})). \dots))),$$

то, применяя последовательно равенство (19), имеем

$$\begin{aligned} \nabla_x f &= \left(\frac{\partial \vec{y}^1}{\partial \vec{x}} \right)^T \cdot \nabla_{y^1} F = \left(\frac{\partial \vec{y}^1}{\partial \vec{x}} \right)^T \cdot \left(\frac{\partial \vec{y}^2}{\partial \vec{y}^1} \right)^T \cdot \nabla_{y^2} F = \dots = \\ &= \left(\frac{\partial \vec{y}^1}{\partial \vec{x}} \right)^T \left(\frac{\partial \vec{y}^2}{\partial \vec{y}^1} \right)^T \cdot \dots \cdot \left(\frac{\partial \vec{y}^N}{\partial \vec{y}^{N-1}} \right)^T \nabla_{y^N} F \end{aligned}$$

(здесь $\nabla_{y^m} F = \nabla_{y^m} F(\vec{y}^N(\vec{y}^{N-1}(\dots \vec{y}^{m+1}(\vec{y}^m). \dots)))$). Таким образом, если функция $f(\vec{x})$ вычисляется как цепочка преобразований

$$\vec{x} \rightarrow \vec{y}^1(\vec{x}) \rightarrow \vec{y}^2(\vec{y}^1) \rightarrow \dots \rightarrow \vec{y}^N(\vec{y}^{N-1}) \rightarrow F(\vec{y}^N), \quad (20)$$

то градиент $\nabla_x f$ может быть вычислен как цепочка линейных преобразований

$$\nabla_{y^N} F \rightarrow \nabla_{y^{N-1}} F \rightarrow \dots \rightarrow \nabla_{y^1} F \rightarrow \nabla_x f, \quad (21)$$

где переход $\nabla_{y^m} F \rightarrow \nabla_{y^{m-1}} F$ осуществляется по формуле

$$\nabla_{y^{m-1}} F = \left(\frac{\partial \vec{y}^m}{\partial \vec{y}^{m-1}} \right)^T \cdot \nabla_{y^m} F. \quad (22)$$

Пусть теперь у нас имеется некоторый алгоритм вычисления функции $f(\vec{x})$, то есть задана последовательность действий:

$$\begin{aligned} x_1 &= x_1 \\ \vdots & \\ x_n &= x_n \\ x_{n+1} &= f_1(x_1, \dots, x_n) \\ x_{n+2} &= f_2(x_1, \dots, x_n, x_{n+1}) \\ \vdots & \\ x_{n+N} &= f_N(x_1, \dots, x_{n+N-1}), \end{aligned} \quad (23)$$

в результате которых получается требуемое значение функции $f(\vec{x}) = z_{n+N}$. Здесь x_1, \dots, x_{n+N-1} — результаты промежуточных действий по вычислению $f(\vec{x})$, $f_j(x_1, \dots, x_{n+j-1})$ — соответствующие операции. В таком случае мы можем представить функцию $f(\vec{x})$ в виде цепочки суперпозиций (20), где

$$\begin{aligned} \vec{y}^1(\vec{x}) &= (y_1^1(\vec{x}), \dots, y_{n+1}^1(\vec{x})) = (x_1, \dots, x_n, f_1(x_1, \dots, x_n)) \\ \vdots & \\ \vec{y}^m(\vec{x}) &= (y_1^m(\vec{y}^{m-1}), \dots, y_{n+m}^m(\vec{y}^{m-1})) = (y_1^{m-1}, \dots, y_{n+m-1}^{m-1}, f_m(y_1^{m-1}, \dots, y_{n+m-1}^{m-1})), \\ \vdots & \end{aligned}$$

то есть вычисление градиента функции $f(\vec{x})$ может быть осуществлено путем выполнения цепочки из N преобразований (21)–(22), причем матрица Якоби имеет в этом случае вид

$$\frac{\partial \vec{y}^m}{\partial \vec{y}^{m-1}} = \begin{pmatrix} 1 & & & & & & \bigoplus \\ & \bigoplus & & & & & \\ & & \dots & & & & \\ & & & \dots & & & \\ & & & & \dots & & \\ & & & & & \dots & \\ \alpha_1^m & \alpha_2^m & \dots & \dots & \dots & \dots & \alpha_{n+m-1}^m \end{pmatrix} \quad (24)$$

где обозначено

$$\alpha_j^m = \frac{\partial}{\partial x_j} f_m(x_1, \dots, x_{n+m-1}); m=1, \dots, N; j=1, \dots, n+m-1$$

Заметим теперь, что если (23) является разложением алгоритма вычисления $f(\vec{x})$ на элементарные операции (операции сложения, вычитания, умножения, деления), то каждая из функций $f_m(x_1, \dots, x_{n+m-1})$ зависит на самом деле не более чем от двух аргументов, то есть среди величин $\{\alpha_j^m\}_{j=1}^{n+m-1}$ только две могут

быть ненулевыми. Это означает, что для вычисления произведения матрицы $(\frac{\partial \bar{y}^m}{\partial \bar{y}^{m-1}})^T$ на вектор $\nabla_{\bar{y}^m} F$ в этом случае достаточно двух операций умножения и двух операций сложения (все компоненты $\nabla_{\bar{y}^m} F$, кроме двух, перейдут в $\nabla_{\bar{y}^{m-1}} F$ без изменений). В свою очередь это означает, что вычислительная стоимость всей цепочки из \mathcal{N} преобразований (21) оценивается $2\mathcal{N}$ операциями умножения. Поскольку разложение (23) означает, что цена вычисления $f(\vec{x})$ — \mathcal{N} операций, то (21)–(22) представляет алгоритм вычисления градиента за цену, пропорциональную цене вычисления функции (при этом коэффициент пропорциональности невелик и не зависит от η).

Поскольку $(\nabla^2 f) \cdot \vec{e} = \nabla(\sum_{k=1}^n \vec{e}_k \frac{\partial f}{\partial x_k})$, то, в силу доказанного общего результата, этот вектор вычисляется так же быстро, как и значение функции $\sum_{k=1}^n \vec{e}_k \frac{\partial f}{\partial x_k}$. Покажем, что это значение вычисляется за ту же цену, что и значение $f(\vec{x})$.

Заметим, что

$$\sum_{k=1}^n e_k \frac{\partial f(\bar{y}(\vec{x}))}{\partial x_k} = \sum_{k=1}^n e_k \sum_{j=1}^m \frac{\partial f}{\partial y_j} \cdot \frac{\partial y_j}{\partial x_k} = \sum_{j=1}^m \left(\sum_{k=1}^n e_k \frac{\partial y_j}{\partial x_k} \right) \frac{\partial f}{\partial y_j},$$

то есть производная по направлению \vec{e} сложной функции $f(\vec{x}) = F(\bar{y}(\vec{x}))$ может быть вычислена как производная функции $F(\bar{y})$ по новому направлению \vec{e}' , определяемому равенством

$$\vec{e}' = \frac{\partial \bar{y}}{\partial \vec{x}} \vec{e},$$

где $\frac{\partial \bar{y}}{\partial \vec{x}}$, как и в (19), — матрица Якоби. Это означает, что если алгоритм вычисления $f(\vec{x})$ представлен цепочкой (20), то производную по направлению можно получить, дифференцируя функцию $F(\bar{y}^{\mathcal{N}})$ по новому направлению $\vec{e}'^{\mathcal{N}}$, которое получается в результате цепочки преобразований

$$\vec{e}' \rightarrow \vec{e}'^2 \rightarrow \dots \rightarrow \vec{e}'^{\mathcal{N}}, \quad (25)$$

где

$$\vec{e}'^m = \left(\frac{\partial \bar{y}^m}{\partial \bar{y}^{m-1}} \right) \vec{e}'^{m-1}. \quad (26)$$

Если теперь мы имеем разложение (23) алгоритма вычисления на элементарные операции, то матрицы $(\frac{\partial \bar{y}^m}{\partial \bar{y}^{m-1}})$ имеют в этом случае вид (24) с двумя ненулевыми элементами в последней строке. Это означает, что преобразование (26) в этом случае требует двух операций сложения и умножения, то есть вычислительная цена всей цепочки преобразований (25) пропорциональна цене вычисления функции $f(\vec{x})$.

2.3. Быстрое дифференцирование суперпозиции функций

Изложенный выше алгоритм быстрого дифференцирования дает нам уверенность в принципиальной возможности вычисления градиента функции за ту же цену, что и вычисление $f(\vec{x})$. Однако в практической ситуации представление в явном виде алгоритма вычисления достаточно сложной функции как суперпозиции элементарных операций может оказаться мало реальным. Мы изложим здесь один прием, позволяющий в случае, когда $f(\vec{x}) = F(\vec{y}(\vec{x}))$, разбить процесс построения алгоритма быстрого дифференцирования для $f(\vec{x})$ на серию задач построения алгоритмов дифференцирования для более простых функций.

Рассмотрим равенство (18) в некоторой фиксированной точке \hat{x} . Пусть

$$\gamma_k = \left. \frac{\partial F(\vec{y})}{\partial y_k} \right|_{\vec{y} = \hat{y}} \quad (27)$$

значения компонент градиента функции $F(\vec{y})$, вычисленные в точке $\hat{y} = \vec{y}(\hat{x})$ (их вычисление, в соответствии с общим результатом раздела 2.2, эквивалентно по сложности переходу $\hat{y} \rightarrow f(\hat{y}) = f(\vec{x})|_{\vec{x} = \hat{x}}$).

Тогда

$$\left. \frac{\partial f}{\partial x_j} \right|_{\vec{x} = \hat{x}} = \sum_{k=1}^m \gamma_k \left. \frac{\partial y_k(\vec{x})}{\partial x_j} \right|_{\vec{x} = \hat{x}} = \left. \frac{\partial}{\partial x_j} \left(\sum_{k=1}^m \gamma_k y_k(\vec{x}) \right) \right|_{\vec{x} = \hat{x}},$$

то есть искомый градиент функции $f(\vec{x})$ может быть определен как градиент вспомогательной функции $\sum_{k=1}^m \gamma_k y_k(\vec{x})$:

$$\nabla_{\vec{x}} f|_{\vec{x} = \hat{x}} = \nabla_{\vec{x}} \left(\sum_{k=1}^m \gamma_k y_k(\vec{x}) \right) \Big|_{\vec{x} = \hat{x}}. \quad (28)$$

Отметим, что стоимость вычисления градиента функции $\sum_{k=1}^m \gamma_k y_k(\vec{x})$ совпадает со стоимостью вычисления градиентов $\nabla_{\vec{x}} y_1(\vec{x}), \nabla_{\vec{x}} y_2(\vec{x}), \dots, \nabla_{\vec{x}} y_m(\vec{x})$ (для доказательства этого факта достаточно заметить, что если в цепочке преобразований (20) $F(\vec{y}^N) = \sum_{k=1}^m \gamma_k y_k^N$, то при построении алгоритма быстрого дифференцирования можно не разлагать далее вычисление на элементарные операции, так как $\nabla_{y^N} F(\vec{y}^N) = (\gamma_1, \gamma_2, \dots)^T$, то есть получается без каких либо вычислений). Таким образом, если построены алгоритмы быстрого дифференцирования для $F(\vec{y}), y_1(\vec{x}), \dots, y_m(\vec{x})$, то вычисление градиента для $f(\vec{x})$ может быть проведено по формулам (27)–(28) за время, пропорциональное времени вычисления $f(\vec{x})$.

Аналогично, производная по направлению может быть рассчитана в два этапа:

$$\delta_j = \sum_{k=1}^n e_k \left. \frac{\partial y_j(\vec{x})}{\partial x_k} \right|_{\vec{x} = \hat{x}}, \quad j=1, 2, \dots, m \quad (29)$$

$$\sum_{k=1}^n e_k \left. \frac{\partial f}{\partial x_k} \right|_{\vec{x} = \hat{x}} = \sum_{j=1}^m \delta_j \left. \frac{\partial F}{\partial y_j} \right|_{\vec{y} = \hat{y}}, \quad (30)$$

общая трудоемкость которых пропорциональна трудоемкости перехода $\vec{x} \rightarrow \vec{y}(\vec{x}) \rightarrow F(\vec{y}(\vec{x}))$.

§ 3. Реализация алгоритмов дифференцирования при уточнении структуры макромолекул

3.1. Быстрое вычисление функционала $R(\vec{x})$

Итак, мы установили, что цена одного шага процедуры (17) минимизации критерия $R(\vec{x})$ определяется ценой расчета величины критерия для текущих значений параметров. Таким образом, проблема быстрого действия программы уточнения структуры свелась к проблеме выбора алгоритма для быстрого вычисления функционала $R(\vec{x})$. Мы будем предполагать, что минимизируемый функционал имеет вид (12). При этом заметим, что, как правило, наиболее трудоемким этапом работы по вычислению значения $R(\vec{x})$ является расчет структурных факторов $\{f_s^R + i f_s^I\}_{s \in S}$, необходимых для определения величины $R_x(\vec{x})$ согласно равенству (1).

Идейно наиболее простой путь вычисления величин $f^R + i f^I$ - это расчет по формуле

$$f_s^R + i f_s^I = \sum_j g_j(s) e^{-\beta_j s^2 / 4} e^{2\pi i (\vec{s}, \vec{r}_j)} \quad (31)$$

где

$$g_j(s) = \frac{2}{S} \int_0^\infty \tau \rho_j^o(\tau) \sin 2\pi s \tau \, d\tau$$

фактор атомного рассеяния атома. Однако за эту простоту придется расплачиваться большим объемом вычислений. Использование свойств симметрии структуры и различные "вычислительные хитрости" позволяют преобразовать формулы (31) к менее трудоемкому виду. Однако, как показал анализ /9/, более существенного выигрыша удастся достичь, применив "обходной" путь расчета величин $f^R + i f^I$. При этом подходе вначале рассчитываются на некоторой сетке u в элементарной ячейке значения функции распределения электронной плотности $\rho(\vec{r})$, а затем определяются величины $f^R + i f^I$ применением обратного преобразования Фурье к функции $\rho(\vec{r})$. Поскольку каждый атом дает существенный вклад в электронную плотность $\rho(\vec{r})$ лишь в небольшом числе узлов вблизи его центра, а на этапе вычисления преобразования Фурье может быть использован алгоритм быстрого преобразования Фурье, то объем вычислений при расчете структурных факторов оказывается практически пропорционален числу атомов и числу структурных факторов.

Использование свойств симметрии при выполнении преобразования Фурье /8/ позволяет рассчитывать функцию $\rho(\vec{r})$ не во всей ячейке, а лишь в некоторой независимой части элементарной ячейки, что также сокращает трудоемкость работы.

Следует обратить внимание, что, вычисляя указанным образом структурные факторы, мы получаем величины, отличающиеся от точных значений. Погрешность возникает из-за того, что генерируя плотность $\rho(\vec{r})$, мы полагаем, что атомы имеют конечный радиус, и того, что выполняется дискретное преобразование Фурье на сетке с конечным шагом. Поэтому шаг сетки и радиусы атомов должны быть выбраны так, чтобы обеспечить точность расчета величин $f^R + i f^I$ не хуже точности экспериментального определения их модулей.

Отметим еще, что использование в критерии (1) величин $f^R + i f^I$, вычисленных описанным образом, означает, что мы несколько изменили минимизируемый критерий, и что, в частности, теперь каждая из величин f_s^R, f_s^I зависит только от конечного числа значений функции $\rho(\vec{r}, \vec{q})$ в узлах сетки u .

Мы не будем в этой работе останавливаться на вычислении функционала $R_s(\vec{X})$; поскольку это более легкая задача, нежели вычисление функционала $R_X(\vec{X})$.

3.2. Вычисление градиента

В этом разделе мы покажем, как изложенный в разделах 2.2 и 2.3 общий алгоритм быстрого дифференцирования реализуется при расчете градиента $\nabla_X R_X(\vec{X})$. Функционал R_X вычисляется согласно 3.1 как цепочка суперпозиций:

$$\vec{q} = \vec{q}(\vec{X}) \quad \text{— обобщенные параметры,} \quad (32)$$

$$\rho_{\vec{r}} = \rho(\vec{r}, \vec{q}) = \sum_j \rho_j(\vec{r}, \vec{q}_j) \quad \text{— значения электронной} \quad (33)$$

$$f_s^R + i f_s^I = c \sum_{\vec{r} \in u} \rho_{\vec{r}} e^{2\pi i (\vec{s}, \vec{r})} \quad \text{— структурные факторы, } \vec{s} \in S \quad (34)$$

$$R_X = \sum_{\vec{s} \in S} a(f_s^R, f_s^I; \vec{s}). \quad (35)$$

В соответствии с общей схемой мы начинаем с вычисления градиента $\{F_s^R, F_s^I\}_{\vec{s} \in S}$ функционала R_X по переменным f_s^R, f_s^I которое не представляет трудности

$$F_s^R = \frac{\partial}{\partial f_s^R} R_X = a_1(f_s^R, f_s^I; \vec{s}), \quad F_s^I = \frac{\partial}{\partial f_s^I} R_X = a_2(f_s^R, f_s^I; \vec{s}), \quad (36)$$

где обозначено $a_1(u, v; \vec{s}) = \frac{\partial}{\partial u} a(u, v; \vec{s}), a_2(u, v; \vec{s}) = \frac{\partial}{\partial v} a(u, v; \vec{s})$.

Далее, воспользовавшись равенством (28), можем определить компоненты $\rho_{\vec{r}}$ градиента функционала R_X по переменным $\rho_{\vec{r}} (\vec{r} \in u)$:

$$\rho_{\vec{r}} = \frac{\partial}{\partial \rho_{\vec{r}}} \left\{ \sum_{\vec{s} \in S} [F_s^R \cdot f_s^R(\rho) + F_s^I \cdot f_s^I(\rho)] \right\} = \quad (37)$$

$$= c \sum_{\vec{s} \in S} (F_s^R \cos 2\pi (\vec{s}, \vec{r}) + F_s^I \sin 2\pi (\vec{s}, \vec{r})) = c \sum_{\vec{s} \in S} (F_s^R + i F_s^I) e^{-2\pi i (\vec{s}, \vec{r})},$$

где при выводе последнего равенства были использованы условие (7) и равенства $F_{\vec{z}}^R = F_{\vec{z}}^R, F_{\vec{z}}^L = -F_{\vec{z}}^L$, вытекающие из условий (8). Как и при выполнении преобразования (34), для расчета $P_{\vec{z}}$ может быть применен алгоритм быстрого преобразования Фурье.

Переход согласно (28) к градиенту $Q = \{Q_{j,\alpha}\}_{j,\alpha}$ по переменным \vec{q} дает:

$$Q_{j,\alpha} = \frac{\partial}{\partial q_{j,\alpha}} \left\{ \sum_{\vec{z} \in \mathcal{U}} P_{\vec{z}} \cdot \rho_{\vec{z}}(\vec{q}) \right\} = \sum_{\vec{z} \in \mathcal{U}} P_{\vec{z}} \frac{\partial}{\partial q_{j,\alpha}} \rho_{\vec{z}}(\vec{z}, \vec{q}_j). \quad (38)$$

Отметим, что в последнем равенстве, как и в равенстве (33), суммирование идет только по точкам $\vec{z} \in \mathcal{U}$, в которых $\rho_{\vec{z}}(\vec{z}, \vec{q}_j) \neq 0$, то есть число операций пропорционально числу атомов.

Наконец, мы можем получить значение градиента $\vec{\chi} = \{\chi_{\kappa,\beta}\}$ по обобщенным параметрам $\vec{\chi}$:

$$\chi_{\kappa,\beta} = \frac{\partial}{\partial \chi_{\kappa,\beta}} \left\{ \sum_{j,\alpha} Q_{j,\alpha} \rho_{j,\alpha}(\vec{\chi}) \right\} = \sum_{j,\alpha} Q_{j,\alpha} \frac{\partial}{\partial \chi_{\kappa,\beta}} \rho_{j,\alpha}(\vec{\chi}). \quad (39)$$

Мы не конкретизируем последнее преобразование, так как у нас не задан конкретный вид зависимостей $\vec{q}(\vec{\chi})$. Отметим только, что из результатов раздела 2.3 следует, что трудоемкость последнего преобразования пропорциональна трудоемкости перехода (32).

Отметим еще, что наличие кристаллографической симметрии позволяет ускорить преобразования (34) и (37).

3.3. Расчет произведения $\nabla_{\vec{\chi}}^2 R_{\chi} \cdot \vec{\omega}$

Пусть задано некоторое направление $\vec{\omega} = \{\omega_{\kappa,\beta}\}_{\kappa,\beta}$ в пространстве обобщенных параметров $\vec{\chi} = \{\chi_{\kappa,\beta}\}_{\kappa,\beta}$. Построенную в предыдущем разделе последовательность преобразований

$$\{\chi_{\kappa,\beta}\}_{\kappa,\beta} \rightarrow \{q_{j,\alpha}\}_{j,\alpha} \rightarrow \{\rho_{\vec{z}}\}_{\vec{z}} \rightarrow \left\{ \begin{matrix} \rho_{\vec{z}}^R \\ \rho_{\vec{z}}^L \end{matrix} \right\}_{\vec{z}} \rightarrow \left\{ \begin{matrix} F_{\vec{z}}^R \\ F_{\vec{z}}^L \end{matrix} \right\}_{\vec{z}} \rightarrow \{\rho_{\vec{z}}\}_{\vec{z}} \rightarrow \{Q_{j,\alpha}\}_{j,\alpha} \rightarrow \{\chi_{\kappa,\beta}\}_{\kappa,\beta} \quad (40)$$

мы можем рассматривать как цепочку суперпозиций (20) при вычислении набора сложных функций $\chi_{\kappa,\beta} = \frac{\partial}{\partial \chi_{\kappa,\beta}} R_{\chi}(\vec{\chi})$. Поскольку набор интересующих нас величин

$$\Omega_{\kappa,\beta} = \sum_{\zeta,\delta} \omega_{\zeta,\delta} \frac{\partial^2}{\partial \chi_{\kappa,\beta} \partial \chi_{\zeta,\delta}} R_{\chi}(\vec{\chi}) = \sum_{\zeta,\delta} \omega_{\zeta,\delta} \frac{\partial}{\partial \chi_{\zeta,\delta}} \chi_{\kappa,\beta}(\vec{\chi})$$

есть не что иное, как совокупность производных по направлению $\vec{\omega}$ от функций $\chi_{\kappa,\beta}(\vec{\chi})$, то, согласно результатам раздела 2.2, для быстрого вычисления величин $\Omega_{\kappa,\beta}$ нам надо построить цепочку направлений (25) в новых переменных, следуя цепочке преобразований (40). При этом для построения новых направлений могут быть использованы равенства (29)-(31).

Мы начинаем с перехода к направлению $\vec{\ell} = \{\ell_{j,\alpha}\}_{j,\alpha}$ в

переменных $\vec{q} = \{q_{j,\alpha}\}_{j,\alpha}$:

$$l_{j,\alpha} = \sum_{k,\beta} \omega_{k,\beta} \cdot \frac{\partial}{\partial x_{k,\beta}} q_{j,\alpha}(\vec{x}). \quad (41)$$

Далее, мы строим направление $\vec{\tau} = \{\tau_{\vec{z}}\}_{\vec{z}}$ в переменных

$$\vec{\rho} = \{\rho_{\vec{z}}\}_{\vec{z}}$$

$$\tau_{\vec{z}} = \sum_{j,\alpha} l_{j,\alpha} \frac{\partial}{\partial q_{j,\alpha}} \rho_{\vec{z}}(\vec{q}) = \sum_j \frac{\partial}{\partial \vec{z}} \rho_j(\vec{z}, \vec{q}_j), \quad (42)$$

то есть функция $\{\tau_{\vec{z}}\}_{\vec{z} \in U}$ есть модифицированная электронная плотность, образованная объединением вкладов производных по направлению от вкладов отдельных атомов в функцию $\rho(\vec{z})$ согласно (33).

Переход к направлению $(\vec{q}^R, \vec{q}^I) = \{(q_S^R, q_S^I)\}_S$ в переменных $(\vec{f}^R, \vec{f}^I) = \{(f_S^R, f_S^I)\}_S$ дается равенствами

$$q_S^R = \sum_{\vec{z}} \tau_{\vec{z}} \frac{\partial}{\partial \rho_{\vec{z}}} f_S^R(\vec{\rho}) = c \sum_{\vec{z} \in U} \tau_{\vec{z}} \cos 2\pi(\vec{s}, \vec{z})$$

$$q_S^I = \sum_{\vec{z}} \tau_{\vec{z}} \frac{\partial}{\partial \rho_{\vec{z}}} f_S^I(\vec{\rho}) = c \sum_{\vec{z} \in U} \tau_{\vec{z}} \sin 2\pi(\vec{s}, \vec{z}),$$

то есть

$$q_S^R + i q_S^I = c \sum_{\vec{z}} \tau_{\vec{z}} e^{2\pi i(\vec{s}, \vec{z})} \quad (43)$$

есть не что иное, как структурные факторы модифицированной плотности $\{\tau_{\vec{z}}\}_{\vec{z}}$.

Далее мы должны перейти к направлению $(\vec{G}^R, \vec{G}^I) = \{(G_S^R, G_S^I)\}_S$ в переменных $(\vec{F}^R, \vec{F}^I) = \{(F_S^R, F_S^I)\}_S$:

$$G_S^R = \sum_{\vec{s}' \in S} [q_{\vec{s}'}^R \cdot \frac{\partial}{\partial f_{\vec{s}'}^R} F_S^R(\vec{f}^R, \vec{f}^I) + q_{\vec{s}'}^I \cdot \frac{\partial}{\partial f_{\vec{s}'}^I} F_S^R(\vec{f}^R, \vec{f}^I)] = \quad (44)$$

$$= q_{\vec{s}'}^R \cdot a_{11}(\vec{f}_{\vec{s}'}^R, \vec{f}_{\vec{s}'}^I; \vec{s}) + q_{\vec{s}'}^I \cdot a_{12}(\vec{f}_{\vec{s}'}^R, \vec{f}_{\vec{s}'}^I; \vec{s})$$

$$G_S^I = \sum_{\vec{s}' \in S} [q_{\vec{s}'}^R \cdot \frac{\partial}{\partial f_{\vec{s}'}^R} F_S^I(\vec{f}^R, \vec{f}^I) + q_{\vec{s}'}^I \cdot \frac{\partial}{\partial f_{\vec{s}'}^I} F_S^I(\vec{f}^R, \vec{f}^I)] =$$

$$= q_{\vec{s}'}^R \cdot a_{21}(\vec{f}_{\vec{s}'}^R, \vec{f}_{\vec{s}'}^I; \vec{s}) + q_{\vec{s}'}^I \cdot a_{22}(\vec{f}_{\vec{s}'}^R, \vec{f}_{\vec{s}'}^I; \vec{s}),$$

где использованы равенства (36) и обозначено $a_{11}(u, v; \vec{s}) = \frac{\partial^2}{\partial u^2} a(u, v; \vec{s})$, $a_{12}(u, v; \vec{s}) = a_{21}(u, v; \vec{s}) = \frac{\partial^2}{\partial u \partial v} a(u, v; \vec{s})$,

$$a_{22}(u, v; \vec{s}) = \frac{\partial^2}{\partial v^2} a(u, v; \vec{s}).$$

Отметим, что в силу общей теории раздела 2.2 даже в случае произвольного функционала $R\chi$ (не обязательно имеющего вид (35)) переход $\vec{f}^R, \vec{f}^I, \vec{q}^R, \vec{q}^I \rightarrow \vec{F}^R, \vec{F}^I, \vec{G}^R, \vec{G}^I$

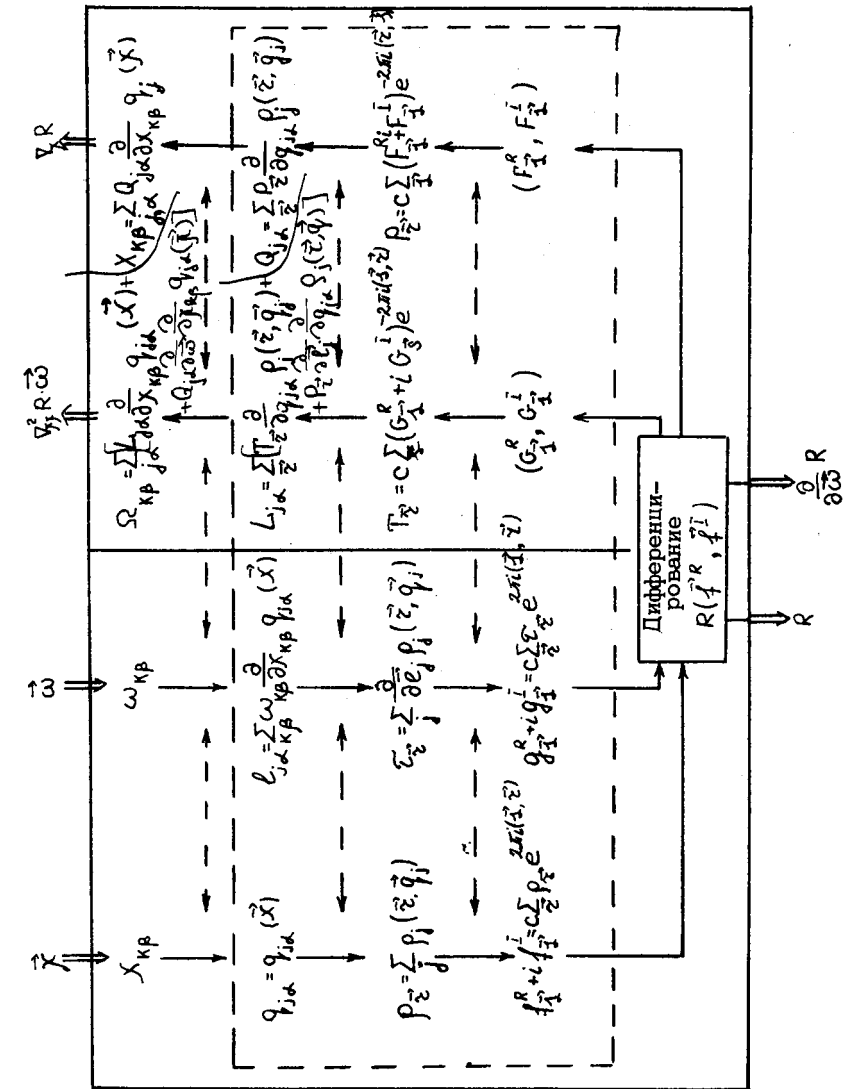


Рис. Схема быстрого вычисления $R_x(\vec{x})$, $\frac{\partial}{\partial \vec{\omega}} R_x$, $\nabla_x R_x$ и $\nabla_{xx}^2 R_x \cdot \vec{\omega}$. Сплошные стрелки показывают последовательность действия. Пунктирные стрелки связывают эквивалентные по времени выполнения операции

не более трудоемко, чем переход от \vec{f}^R, \vec{f}^I к $R_x(\vec{f}^R, \vec{f}^I)$.

Используя равенство (37), мы получаем для направления $\vec{T} = \{T_{\vec{z}}\}_{\vec{z}}$ в переменных $\vec{P} = \{P_{\vec{z}}\}_{\vec{z}}$ выражение

$$T_{\vec{z}} = \sum_{\vec{s} \in S} [G_{\vec{s}}^R \cdot \frac{\partial}{\partial P_{\vec{s}}} P_{\vec{z}}(\vec{f}^R, \vec{f}^I) + G_{\vec{s}}^I \cdot \frac{\partial}{\partial P_{\vec{s}}} P_{\vec{z}}(\vec{f}^R, \vec{f}^I)] =$$

$$= c \sum_{\vec{s} \in S} (G_{\vec{s}}^R + i G_{\vec{s}}^I) e^{-2\pi i(\vec{s}, \vec{z})} \quad (45)$$

и, вычисляя последовательно направления в переменных \vec{Q} и \vec{X} , получаем

$$L_{j,\alpha} = \sum_{\vec{z} \in U} T_{\vec{z}} \cdot \frac{\partial}{\partial P_{\vec{z}}} Q_{j,\alpha}(\vec{P}) + \sum_{\vec{z} \in U} P_{\vec{z}} \frac{\partial}{\partial q_j} \frac{\partial}{\partial P_{\vec{z}}} Q_{j,\alpha}(\vec{z}, \vec{q}_j) \quad (46)$$

$$\Omega_{\kappa,\beta} = \sum_{j,\alpha} L_{j,\alpha} \cdot \frac{\partial}{\partial Q_{j,\alpha}} X_{\kappa,\beta}(\vec{Q}) + \sum_{j,\alpha} Q_{j,\alpha} \frac{\partial}{\partial X_{\kappa,\beta}} \frac{\partial}{\partial Q_{j,\alpha}} X_{\kappa,\beta}(\vec{X}). \quad (47)$$

Отметим полную аналогичность соответствующих этапов расчета величин $X_{\kappa,\beta}$ по формулам (32)-(39) и величин $\Omega_{\kappa,\beta}$ по формулам (41)-(47).

3.4. Резюме

Итак, мы показали, что при любом способе описания атомной модели обобщенными параметрами \vec{X} и произвольном виде минимизируемого функционала $R(\vec{X})$ может быть построен алгоритм, позволяющий рассчитать значения функции $R(\vec{X})$ и производной по направлению $\frac{\partial}{\partial \vec{\omega}} R(\vec{X})$ и всех компонент векторов $\nabla_x R$ и $\nabla_{xx}^2 R \cdot \vec{\omega}$ за время, равное учетверенному времени расчета одного значения $R(\vec{X})$. На рисунке приведена схема построения такого алгоритма. Отметим, что центральная часть алгоритма (обведенная на рисунке пунктиром) не зависит от способа введения параметров \vec{X} и вида функционала $R(\vec{f}^R, \vec{f}^I)$ и может быть реализована универсальным образом. Привязка алгоритма к конкретному описанию модели и критерию уточнения заключается лишь в конкретизации переходов

$$\vec{X} \rightarrow q(\vec{X}) \text{ и } \{\vec{f}^R, \vec{f}^I\} \rightarrow R.$$

Отметим также, что для критериев, выражающихся через параметры атомов (критерия R_S), принципиальная схема расчета аналогична, но имеет меньшую "глубину" - после определения величин \vec{q} мы сразу уходим на вычисление критерия, а не выполняем еще преобразования $\vec{q} \rightarrow \vec{p} \rightarrow (\vec{f}^R, \vec{f}^I)$.

ВЫВОДЫ

1. Способ уточнения атомной модели макромолекулы определяется двумя факторами:

а) выбором обобщенных независимых параметров для описания текущего состояния структуры;

б) выбором критерия соответствия модели экспериментальным данным.

2. Жесткие ограничения на параметры модели могут быть учтены путем надлежащего выбора обобщенных параметров. Не жесткие ограничения вводятся путем добавления соответствующих слагаемых в критерий, минимизируемый в процессе уточнения.

3. В критерий соответствия модели рентгеновскому эксперименту целесообразно включать фазовую информацию, причем это включение должно учитывать надежность фазовой информации. Примером такого критерия является критерий (5).

4. Техника быстрого дифференцирования позволяет реализовать метод сопряженных градиентов для минимизации критерия уточнения таким образом, что стоимость одного шага уточнения определяется только стоимостью вычисления значения критерия для текущего состояния структуры. При этом для атомных моделей может быть предложен единый алгоритм такого рода.

5. Использование алгоритма быстрого преобразования Фурье позволяет реализовать вычисление критерия (а тем самым и цикл уточнения) за время, линейно зависящее от размеров молекулы.

Авторы благодарны К.В.Киму и Ю.Е.Нестерову, изложившим авторам алгоритм быстрого дифференцирования, оказавший существенное идейное влияние на данную работу.

ЛИТЕРАТУРА

1. Agarwal R.C. - Acta Cryst., 1978, A34, p. 791-809.
2. Booth A.D. - Proc. Roy. Soc; London Ser. A, 1949, No. 1048, v. 197, p. 336-355.
3. Diamond R. - Acta Cryst., 1971, v. A27, p. 436-452.
4. Dodson E.J., Isaacs N.W., Rollett J.S. - Acta Cryst., 1976, v. A32, p. 311-315.
5. Hendrickson W.A., Lattman E.E. - Acta Cryst., 1970, v. B26, p. 136-143.
6. Jack A., Levitt M. - Acta Cryst., 1978, v. A34, p. 931-935.
7. Scheringer C. - Acta Cryst., 1982, v. A38, p. 618-623.
8. Ten Eyck L.F. - Acta Cryst., 1973, v. A29, p. 183-191.
9. Ten Eyck L.F. - Acta Cryst., 1977, v. A33, p. 486-492.
10. Ким К.В., Нестеров Ю.Е., Скоков В.А., Черкасский Б.В. - В сб.: Тезисы докладов II конференции "Оптимальное планирование и управление народным хозяйством". ЦЭМИ АН СССР, М., 1983, с. 73-76.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

§ 1. Постановка задачи уточнения модели	4
1.1. Модель исследуемого объекта	4
1.2. Уточнение модели по данным рентгеновского рас- сеяния	5
1.3. Ограничения на параметры атомов	7
1.4. Резюме	8
§ 2. Минимизация функций большого числа переменных	9
2.1. Метод сопряженных градиентов	9
2.2. Алгоритм быстрого дифференцирования	11
2.3. Быстрое дифференцирование суперпозиции функций	15
§ 3. Реализация алгоритмов дифференцирования при уточ- нении структуры макромолекул	16
3.1. Быстрое вычисление функционала $R(\vec{x})$	16
3.2. Вычисление градиента	17
3.3. Расчет произведения $\nabla_{xx}^2 R_{\chi} \cdot \vec{\omega}$	18
3.4. Резюме	21
ВЫВОДЫ	22
ЛИТЕРАТУРА	22

Владимир Юрьевич Лукин, Александр Георгиевич Уржумцев

КОНСТРУИРОВАНИЕ ПРОГРАММ ДЛЯ УТОЧНЕНИЯ
АТОМНОЙ СТРУКТУРЫ МАКРОМОЛЕКУЛ НА БАЗЕ
АЛГОРИТМОВ БЫСТРОГО ПРЕОБРАЗОВАНИЯ ФУРЬЕ
И БЫСТРОГО ДИФФЕРЕНЦИРОВАНИЯ

Препринт

Отредактировано и подготовлено к печати в ОНТИ НЦБИ

Редактор С.Я.Гудкова

Технический редактор С.М.Ткачук

Корректоры Т.К.Крейцер, Л.М.Орлова

Подписано в печать 24.10.83 г. Т-20175. Уч.-изд.л. 1,3.

Тираж 150 экз. Заказ 4004Р. Бумага офсетная.

Бесплатно. Формат 60х90/16. Изд. № 260.

Отпечатано на ротапринте в Отделе научно-технической
информации Научного центра биологических исследований АН СССР
в Пущине