

УДК 544.424.2

## ОБЩИЙ ПОДХОД К ОЦЕНКЕ ТОЧНОСТИ МЕТОДОВ МОЛЕКУЛЯРНОЙ РЕКОНСТРУКЦИИ СЛОЖНЫХ СМЕСЕЙ

© 2024 г. Н. А. Глазов<sup>a</sup>, \*, А. Н. Загоруйко<sup>a</sup><sup>a</sup>ФГБУН Институт катализа им. Г.К. Борескова СО РАН,  
просп. Акад. Лаврентьева, 5, Новосибирск, 630090 Россия

\*e-mail: glazov@catalysis.ru

Поступила в редакцию 30.08.2024 г.

После доработки 09.12.2024 г.

Принята к публикации 10.12.2024 г.

В работе обобщен ряд методов молекулярной реконструкции и предложен общий подход к определению надежности реконструированного состава, учитывающий точность измерений. Рассматриваемая процедура оценивает ковариационную матрицу параметров модели на основе матрицы производных расчетных свойств по параметрам модели и ковариационной матрицы измерений, что позволяет определить диапазон, в котором могут варьироваться концентрации компонентов, не противореча имеющимся экспериментальным данным. Представленная процедура достаточно проста в применении при вычислении производных с использованием прямого автоматического дифференцирования.

**Ключевые слова:** математическое моделирование, молекулярная реконструкция, максимум энтропии, стохастическая реконструкция, погрешность

DOI: 10.31857/S0453881124050044, EDN: QVSCRG

### ВВЕДЕНИЕ

Моделирование кинетики химических реакций давно стало неотъемлемой частью как теоретического познания мира, так и практических приложений.

Классическое описание химических реакций через константы скорости, порядки реакций и энергии активации предоставляет хорошую теоретическую базу для моделирования огромного количества химических процессов, однако с увеличением числа протекающих реакций в системе количество определяемых параметров растет пропорционально, и для систем из десятков тысяч компонент объем необходимых экспериментов для нахождения всех параметров для каждой реакции становится просто неподъемным.

**Сокращения и обозначения:**  $x_i$  – мольная доля  $i$ -го компонента;  $H$  – информационная энтропия Шеннона;  $M$  – молекулярная масса;  $\omega$  – массовая доля углерода;  $J$  – Якобиан (матрица производных расчетных свойств по параметрам модели);  $A, B, C, D$  – элементы матрицы Якобиана  $2 \times 2$  в примерах;  $E$  – математическое ожидание;  $N$  – количество соединений в смеси;  $\text{Var}$  – дисперсия;  $p$  – свойство индивидуального компонента;  $P$  – вероятность;  $\sigma$  – среднеквадратичное отклонение;  $\bar{X}$  – среднее значение величины  $X$ ;  $X^T$  – транспонированная матрица  $X$ ;  $X'$  – производная  $X$  по своему аргументу;  $X'_p$  – производная  $X$  по  $p$  для функций многих переменных.

В связи с этим для моделирования, например, реакций в тяжелых нефтяных фракциях вводятся дополнительные предположения, позволяющие сократить число определяемых параметров. Достаточно популярный подход заключается в “агрегировании” разных соединений в групповые компоненты. Тем самым ставится знак равенства между их кинетическими параметрами, что можно рассматривать как достаточно простой способ моделировать эти параметры. Например, в работе [1] именно таким образом решают проблему большого количества компонентов при моделировании каталитического риформинга бензина, а в работе [2] биомассу представляют в виде трехкомпонентной смеси.

В литературе известны более сложные модели для кинетических констант. Например, в работе [3] авторы описали 146 реакций с помощью всего 22 экспериментально определяемых параметров, выражая энергию активации однотипных реакций линейной функцией от термодинамических характеристик компонентов.

Идея моделировать значения, традиционно находимые экспериментально, распространяется не только на кинетические параметры, но и на концентрации компонентов, что особенно полезно при применении детальных кинети-

ческих моделей [4]. Такие методы в литературе часто называют методами “молекулярной реконструкции” [5], и уже сейчас существует достаточно много разных вариантов, отличающихся предположениями при построении состава, способами определения параметров и тем, что понимается под “компонентом” смеси (начиная от молекул, псевдокомпонентов и узких фракций и заканчивая другими сложными смесями).

Однако, как и в случае любых других моделей, возникают вопросы о точности и адекватности моделей реконструкции. Цель настоящей работы – обобщение нескольких подходов к молекулярной реконструкции и установление общих закономерностей влияния экспериментальных данных на точность реконструкции.

## ТЕОРЕТИЧЕСКАЯ ЧАСТЬ

### *Обобщенный взгляд на молекулярную реконструкцию*

Как отмечалось во введении, ключевым предположением молекулярной реконструкции является существование функциональной связи между (неизвестными) концентрациями и (известными) характеристиками компонентов:

$$x_i = \tilde{f}(p_i), \quad (1)$$

где  $x_i$  – концентрация  $i$ -го компонента,  $p_i$  – свойство  $i$ -го компонента, а  $\tilde{f}$  – функция.

Что можно сказать о свойствах такой функции? Поскольку сумма всех концентраций должна равняться единице, то

$$\tilde{f}(p_i) = x_i = \frac{x_i}{1} = \frac{x_i}{\sum x_j} = \frac{f(p_i)}{\sum f(p_j)}, \quad (2)$$

где  $f$  – функция, описывающая концентрации с точностью до постоянного множителя (в отличие от  $\tilde{f}$ ). Такая функция должна быть одного знака и не равняться нулю сразу во всех рассматриваемых точках.

Разные методы реконструкции получаются при разных стратегиях выбора  $f(x)$ . Например, если все концентрации ненулевые, то достаточно естественно искать решение в форме  $e^{g(x)}$ , что часто делается в методе реконструкции максимизации энтропии [6]. Стохастическая реконструкция [7, 8] получается, если задать  $f(x) = F(x) - F(x - 1)$  для кумулятивного распределения  $F$  непрерывной случайной величины (часто гамма-распределения или экспоненциального).

В случае, если в распоряжении имеются данные о нескольких характеристиках, то  $f(p_1, p_2, \dots, p_n)$  часто представляют в виде произведения функций одной переменной, то есть:

$$f(p_1, p_2, \dots, p_n) = f_1(p_1) \times f_2(p_2) \times \dots \times f_n(p_n). \quad (3)$$

Причем каждая функция включает в себя как минимум один параметр, при варьировании которого можно получать смеси с разными свойствами. При реконструкции образца происходит оценка параметров модели с помощью методов оптимизации. Это дает возможность использовать абсолютно любые данные, которые возможно оценить, зная детальный состав, что дает невообразимую свободу выбора экспериментальных данных, однако это же вносит и много неопределенности в надежность находимого состава.

Молекулярная реконструкция позволяет найти состав (или составы), который не противоречит нашим знаниям о качественном составе и собранным экспериментальным данным, но вышесказанное не означает, что реконструированная смесь обязана хорошо имитировать все химические свойства образца.

Это легко увидеть, если довести ситуацию до абсурда, попытавшись реконструировать образец вообще без каких-либо данных. Абсолютно любой состав в этом случае естественно не противоречит тому, что мы знаем о смеси, а значит, является адекватным решением задачи о реконструкции при имеющейся информации, но сложно ожидать какой-то точности при описании кинетики с таким составом.

Нередки ситуации, когда ряд структурных характеристик, сильно сказывающихся на реакционной способности смеси, слабо влияет на находимые в эксперименте свойства, что может привести к тому, что по экспериментальным данным не удастся определить эти самые структурные особенности, или, в контексте оптимизационных задач, любое (даже ничем не обоснованное) значение не будет противоречить эксперименту.

В связи с вышесказанным представляется, что для использования реконструированного состава для моделирования кинетики необходимо обеспечить зависимость параметров модели от расчетных свойств.

### *Предлагаемый метод оценки точности параметров модели*

Оценки расчетных свойств находятся однозначно исходя из набора концентраций, а кон-

центрации в свою очередь обычно однозначно определяются на основе параметров модели. Тем самым можно установить в соответствие расчетные свойства смеси и набор параметров модели:

$$G(u) = m, \quad (4)$$

где  $G(u)$  – функция, ставящая в соответствие параметры модели и расчетные свойства, а  $m$  – экспериментально измеренные свойства образца.

Реконструировать образец – значит найти такое  $u$ , чтобы  $G(u)$  равнялось заданному экспериментальному набору свойств. Стоит отметить, что измерения ( $m$ ) в реальности заданы с какой-то погрешностью  $\Delta m$ , а значит, нужно ожидать, что и параметры модели ( $u$ ) могут содержать неопределенность ( $\Delta u$ ):

$$G(u + \Delta u) = m + \Delta m. \quad (5)$$

Если считать, что отклонения невелики, то возможно разложить зависимость расчетных

свойств от параметров модели в ряд Тейлора в окрестности  $u_0$ :

$$G(u_0) + J\Delta p = m + \Delta m, \quad (6)$$

где  $J$  – матрица Якоби.

При реконструировании образца обычно находится такие  $u_0$ , что  $G(u_0) = m$ , а значит  $J\Delta u = \Delta m$ . Стоит отметить, что  $\Delta m$  – достаточный понятный случайный вектор. В отсутствие систематической ошибки можно ожидать, что каждый элемент этого вектора описывает случайную ошибку с нулевым математическим ожиданием. Более того, в силу независимости измерений различных свойств можно считать, что недиагональные элементы ковариационной матрицы равны нулю. Диагональные элементы являются квадратами среднеквадратичных отклонений и обычно известны из методик.

Используя определение ковариационной матрицы, можно получить связь между ковариационной матрицей параметров модели и ковариационной матрицей измерений:

$$\text{cov}(\Delta m) = E\left((\Delta m - E(\Delta m))(\Delta m - E(\Delta m))^T\right) = E(\Delta m \Delta m^T), \quad (7)$$

$$E(\Delta m \Delta m^T) = E(J \Delta u \Delta u^T J^T) = J \text{cov}(\Delta u) J^T. \quad (8)$$

Для обратимого якобиана возможно получить достаточно простое выражение для ковариационной матрицы параметров:

$$\text{cov}(\Delta u) = J^{-1} \text{cov}(\Delta m) J^{-1T}. \quad (9)$$

Если якобиан необратим, то, вероятно, наилучшей оценкой ковариационной матрицы параметров может служить аналогичное выражение с псевдообратной матрицей:

$$\text{cov}(\Delta u) = J^+ \text{cov}(\Delta m) J^{+T}. \quad (10)$$

На диагонали этой матрицы можно найти дисперсии для всех параметров, а недиагональные элементы характеризуют взаимосвязь между параметрами.

У полученного выражения существует и альтернативная интерпретация, связанная с линейным преобразованием матрицы Фишера (которой можно считать обратную матрицу к ковариационной матрице измерений).

На практике для вычисления якобиана хорошо подходит прямое автоматическое дифференцирование.

#### Реконструкция максимизацией энтропии

В терминах вышеописанного подхода реконструкция максимизации энтропии ищет  $f(p)$  в виде  $e^{g(p)}$ . Функция  $g(p)$  описывает все налагаемые ограничения на состав, будь то строгие/нестрогие ограничения на взвешенные средние или дисперсии. При наличие большего количества ограничений  $f(p)$ , как отмечалось ранее, перемножаются, а значит  $g(p)$  просто складываются.

Для строгих ограничений на среднее  $g(p) = -\lambda p$ , где  $p$  представляет собой вес, с которым данный компонент входит в взвешенное среднее, а  $\lambda$  – это настраиваемый параметр модели (одинаковый для всех компонентов).

Для нестрогих ограничений на среднее  $g(p) = \varepsilon(p/\sigma)$ , где  $\varepsilon$  – это настраиваемый параметр модели, и  $\sigma$  – это заданное значение среднеквадратичного отклонения, характеризующего то, насколько мы не уверены в точности экспериментально полученного значения для среднего.

Для строгих ограничений на дисперсию  $g(p) = -v(p - p_0)^2$ , где  $v$  – параметр модели, а  $p_0$  – это среднее.

Строгий вывод данных выражений опирается на принцип максимума энтропии, где под энтропией понимается исключительно характеристика распределения, и она никак (по крайней мере явным образом) не связана с энтропией соединений, входящих в смесь:

$$H = -\sum_{i=1}^N x_i \ln x_i. \quad (11)$$

Задача максимизации решается с учетом налагаемых ограничений, с помощью которых включаются экспериментальные данные. Задачи условной оптимизации могут быть превращены в задачи безусловной оптимизации в пространстве большей размерности с использованием метода множителей Лагранжа. Часто в литературе рассматривают строгие и нестрогие ограничения на среднее и среднеквадратичное отклонения. В терминах оптимизационных задач они записываются следующим образом:

$$H = -\sum_{i=1}^N x_i \ln x_i + \mu \left( 1 - \sum_{i=1}^N x_i \right) + \sum_{s=1}^S \lambda_s \left( f_s - \sum_{i=1}^N x_i f_{i,s} \right), \quad (12)$$

$$H = -\sum_{i=1}^N x_i \ln x_i + \mu \left( 1 - \sum_{i=1}^N x_i \right) - \frac{1}{2} \sum_{k=1}^K \frac{\left( f_j - \sum_{i=1}^N x_i f_{i,j} \right)^2}{\sigma^2}, \quad (13)$$

$$H = -\sum_{i=1}^N x_i \ln x_i + \mu \left( 1 - \sum_{i=1}^N x_i \right) + \sum_{t=1}^T \nu_t \left( \sigma_r^2 - \sum_{i=1}^N x_i (f_t - f_{i,t})^2 \right), \quad (14)$$

где  $f$  с одним индексом – это значение свойства для смеси,  $f$  с двумя индексами – это свойства чистых компонентов,  $S$  и  $K$  – это количества строгих и нестрогих ограничений на взвешенное среднее,  $T$  – количество строгих ограничений на среднеквадратичное отклонение, а  $\mu$ ,  $\lambda$ ,  $\nu$  – множители Лагранжа.

Первое слагаемое представляет собой информационную энтропию, второе слагаемое накладывает ограничение на сумму концентраций. Третье слагаемое описывает строгие/нестрогие ограничения на среднее/строгое ограничение на среднеквадратичное отклонение соответственно.

Можно заметить, что в такой форме они все зависят от большего числа переменных, чем количество искомых концентраций. Однако для задач с такими ограничениями (и любыми комбинациями таковых) существует достаточно простая зависимость концентраций компонентов от множителей Лагранжа  $x = e^{g(\lambda, \varepsilon, \nu)} / Z$  ( $Z$  – нормировочный множитель), не включающая другие концентрации. Поэтому возможно полностью избавиться от концентраций в оптими-

зационной задаче, заметно сокращая ее размерность. С учетом этого можно записать:

$$H = \ln Z + \sum_{s=1}^S \lambda_s f_j, \quad (15)$$

$$H = \ln Z + \sum_{k=1}^K \left( \frac{1}{2} \varepsilon_k^2 - \frac{f_k}{\sigma_k} \varepsilon_k \right), \quad (16)$$

$$H = \ln Z + \sum_{t=1}^T \nu_t \sigma_t^2. \quad (17)$$

Как видно, число неизвестных в этом случае просто равняется числу ограничений, что на практике достаточно удобно, когда данных гораздо меньше, чем определяемых концентраций.

В общем случае такой “фокус” не работает, и  $g(p)$  будет включать в себя концентрации компонентов в явном виде, что может превратить задачу в достаточно нетривиальную.

Рассмотрим общий случай, при котором оценка свойства определяется функцией  $L(x)$ :

$$H = -\sum_{i=1}^N x_i \ln x_i + \mu \left( 1 - \sum_{i=1}^N x_i \right) + \sum_{s=1}^S \lambda_s (f_s - L(x)). \quad (18)$$

Частная производная этого выражения по концентрации  $i$ -го компонента:

$$\frac{\partial H}{\partial x_i} = -\ln x_i - 1 - \mu - \sum_{s=1}^S \lambda_s \frac{\partial L}{\partial x_i}. \quad (19)$$

В условиях оптимума производная будет равна нулю, и тогда:

$$-\ln x_i - 1 - \mu - \sum_{s=1}^S \lambda_s \frac{\partial L}{\partial x_i} = 0, \quad (20)$$

$$\ln x_i = -\mu - 1 - \sum_{s=1}^S \lambda_s \frac{\partial L}{\partial x_i}, \quad (21)$$

$$x_i = e^{-\mu-1-\sum_{s=1}^S \lambda_s \frac{\partial L}{\partial x_i}}, \quad (22)$$

$$x_i e^{\mu+1} = e^{-\sum_{s=1}^S \lambda_s \frac{\partial L}{\partial x_i}}. \quad (23)$$

Просуммировав такие выражения для каждого  $i$ :

$$e^{\mu+1} \sum_{i=1}^N x_i = \sum_{i=1}^N e^{-\sum_{s=1}^S \lambda_s \frac{\partial L}{\partial x_i}}. \quad (24)$$

Поскольку сумма концентраций равна 1, то уравнение (22) можно переписать в виде:

$$x_i = \frac{e^{-\sum_{s=1}^S \lambda_s \frac{\partial L}{\partial x_i}}}{e^{\mu+1}} = \frac{e^{-\sum_{s=1}^S \lambda_s \frac{\partial L}{\partial x_i}}}{\sum_{i=1}^N e^{-\sum_{s=1}^S \lambda_s \frac{\partial L}{\partial x_i}}}. \quad (25)$$

В наших прошлых обозначениях это просто значит, что в общем случае для строго ограниченного среднего  $g(x) = -\lambda L'(x)$ , где  $L'$  – это производная правил смешения по концентрации компонента. Если правила смешения сложнее, чем линейные, то производная по концентрации не будет константой, что в конечном счете и позволяло упрощать задачу в рассмотренных выше случаях.

### Стохастическая реконструкция

Другой достаточно популярный подход к моделированию составов представляет компоненты как комбинацию “структурных атрибутов”, которые, как понятно из названия, характеризуют структуру молекул. Концентрации компонентов определяются как функции этих самых структурных атрибутов. Структурные атрибуты могут принимать как числовые значения (длина цепи парафина, количество/длина заместителей, количество колец и т.д.) так и категориальные (тип молекулы).

В качестве основы зависимости концентрации от структурных атрибутов используется кумулятивные распределения случайных величин. Например, можно задавать  $f(x)$  как

$$f(x) = F(x) - F(x-1) \quad (26)$$

для произвольного кумулятивного распределения  $F$ .

Например, если структурный атрибут может принимать значения от 1 до  $Q$ , то вероятность получить  $q$  будет:

$$P(X = q) = \frac{F(q) - F(q-1)}{F(Q) - F(0)}. \quad (26)$$

Что эквивалентно тому, что генератор случайных чисел генерирует случайную величину из распределения  $F$  и округляет ее вверх. Если полученное число находится за пределами области допустимых значений, все начинается сначала. (Знаменатель является результатом суммирования бесконечно убывающей геометрической прогрессии в данной интерпретации.) Впрочем, если генерация выходит за пределы выбранного диапазона достаточно часто, это может свидетельствовать о том, что выбранная модель плохо описывает состав.

В классической реализации стохастической реконструкции составы получаются именно по такой схеме с генерацией случайных чисел. Концентрации определяются как наблюдаемые частоты (т.е. после генерации  $N$  “молекул” таким способом, каждой присваивается концентрация  $1/N$ ).

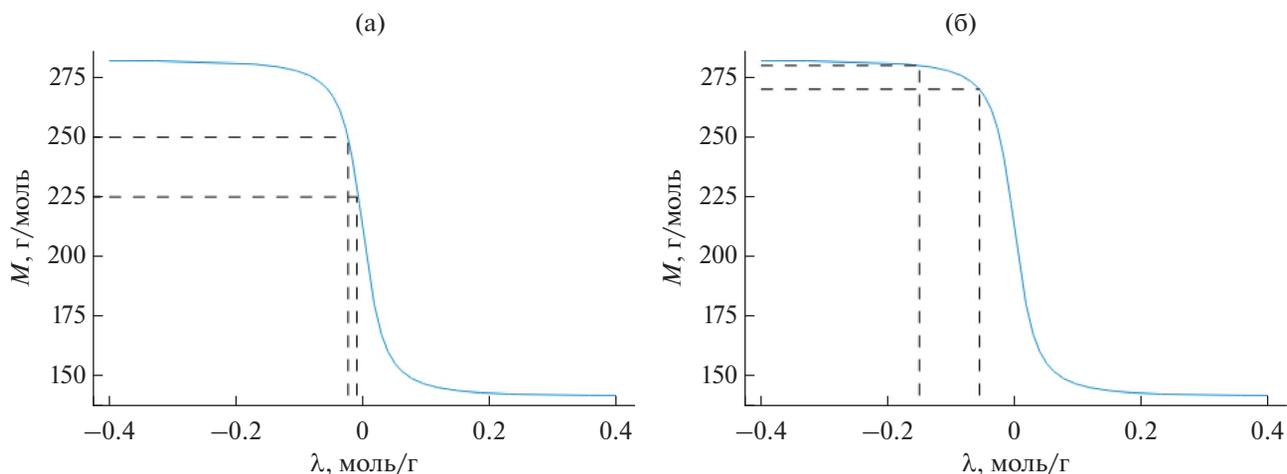
## ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

### Метод максимизации энтропии

Предположим, что имеется смесь нормальных парафинов (для ясности, скажем,  $C_{10}-C_{20}$ ) и известна средняя молекулярная масса этой смеси. Опираясь на эти данные, почти наверняка невозможно определить состав точно, но можно его оценить с помощью метода максимизации энтропии.

Средняя молекулярная масса смеси представляет собой взвешенную сумму молекулярных масс компонентов. Используя строгие ограничения  $f(M) = e^{-\lambda M}$  и варьируя значение  $\lambda$ , можно добиться почти любого значения средней молекулярной массы.

Используя рис. 1, можно понять, что погрешность в молекулярной массе в разной степени



**Рис. 1.** Зависимость расчетной средней молекулярной массы от параметра модели: широкий диапазон молекул масс отвечает узкому диапазону параметров (а) и узкий диапазон молекулярных масс отвечает широкому диапазону параметров (б).

влияет на погрешность в описании параметров. Достаточно широкий диапазон значений молекулярных масс может отвечать очень узкому диапазону параметров и, наоборот, погрешность в определении свойств может соответствовать очень широкому диапазону параметров модели и, как следствие, концентраций компонентов. Так, в вышеприведенном примере, если мы знаем, что средняя молекулярная масса находится между 270 и 280, оценка концентрации  $C_{20}$  в смеси будет варьироваться от 53 до 88%.

Несложно представить ситуацию, в которой скорости реакций тяжелых парафинов будут значительно выше, чем скорости легких. Соответственно, на результаты моделирования силь-

но повлияет то обстоятельство, составляют ли тяжелые парафины 90% всей смеси или 50%.

Предлагаемый метод оценки неопределенности параметров в данном случае утверждает:

$$\sigma_u^2 = \frac{1}{M'(\lambda)} \sigma_M^2 \frac{1}{M'(\lambda)} = \left( \frac{\sigma_M}{M'(\lambda)} \right)^2, \quad (27)$$

$$\sigma_u = \frac{\sigma_M}{|M'(\lambda)|}. \quad (28)$$

Для удобства вычислений можно показать, что производная при таких правилах смешения равна дисперсии с точностью до знака:

$$x_i = \frac{e^{-\lambda M_i}}{\sum e^{-\lambda M}}, \quad (29)$$

$$x_i' = \frac{-M_i e^{-\lambda M_i} \sum e^{-\lambda M} + e^{-\lambda M_i} \sum M e^{-\lambda M}}{(\sum e^{-\lambda M})^2} = -M_i x_i + x_i \bar{M}, \quad (30)$$

$$\left( \sum_{i=1}^N M_i x_i \right)' = -\sum_{i=1}^N M_i^2 x_i + \bar{M}^2 = \bar{M}^2 - \overline{M^2} = -\text{Var}(M). \quad (31)$$

Можно построить зависимость среднеквадратичного отклонения параметра ( $\sigma$ ) от значения параметра (рис. 2), а также среднюю абсолютную ошибку в определении концентраций (рис. 3).

Можно видеть, что чем дальше смесь от эквивалентной, тем выше ошибка определения параметра и концентраций. Таким образом, становится понятна важность работы в областях с низким (по

модулю) параметром, для чего в смесь могут вводиться дубликаты уже существующих компонентов, позволяя добиваться большей точности находимых концентраций. Введение дубликатов помогает, помимо всего прочего, предотвращать переполнение при вычислении экспонент.

При использовании нестрогих ограничений ( $f(x) = e^{cx/\sigma}$ ) состав становится ближе к экви-

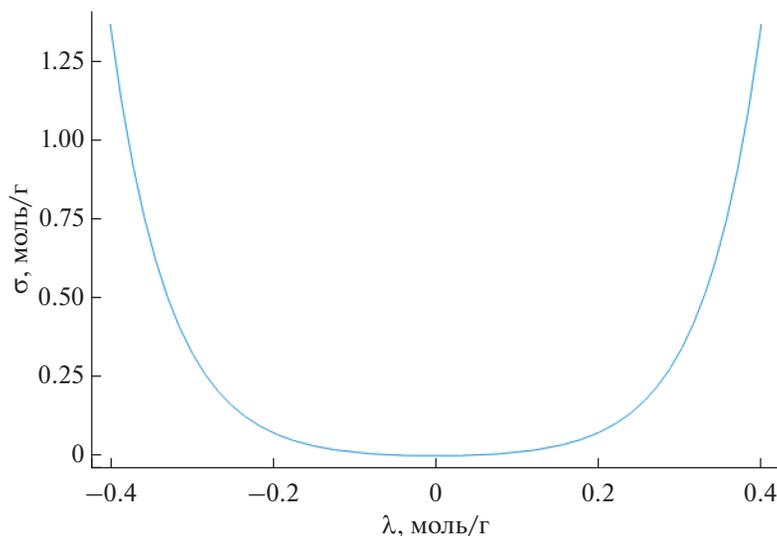


Рис. 2. Зависимость погрешности находимого параметра от значения параметра.

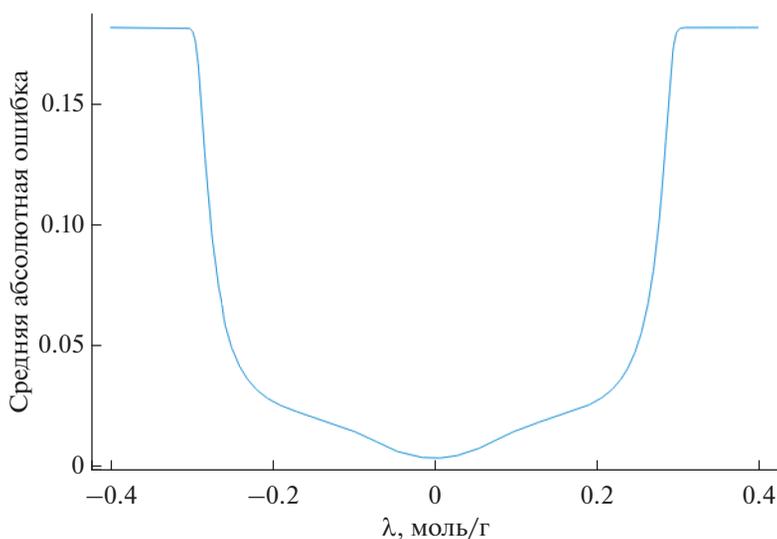


Рис. 3. Зависимость средней абсолютной ошибки определения концентрации от параметра модели.

молярному по сравнению с теми же значениями для строгих ограничений (рис. 4). Можно заметить, что показатель экспоненты в этой модели не содержит  $-1$  из строгих ограничений. Это не имеет принципиального эффекта на решение.

Несложно заметить, что четкие ограничения с параметром  $\lambda$  соответствуют нечетким ограничениями с  $-\varepsilon/\sigma$  (т.е. отличаются на постоянный множитель), а значит, производная молекулярной массы по  $\varepsilon$  будет:

$$M'(\varepsilon) = \frac{dM}{d\lambda} \frac{d\lambda}{d\varepsilon} = \frac{\text{Var}(M)}{\sigma}. \quad (32)$$

Тем самым,

$$\sigma_u^2 = \frac{1}{M'(\mu)} \sigma_M^2 \frac{1}{M'(\mu)} = \frac{\sigma_M^2 \sigma^2}{\text{Var}(M)^2}, \quad (32)$$

$$\sigma_u = \frac{\sigma \sigma_M}{\text{Var}(M)}. \quad (33)$$

В сущности это означает, что неопределенности перемножаются.

Введем в рассмотрение дополнительные измерения, скажем, среднюю долю углерода в смеси. Обычно,  $f(M, \omega) = f_1(M) \times f_2(\omega) = e^{-\lambda_1 M - \lambda_2 \omega}$ . Вычислим Якобиан. Для начала определимся с производными концентраций по параметрам модели:

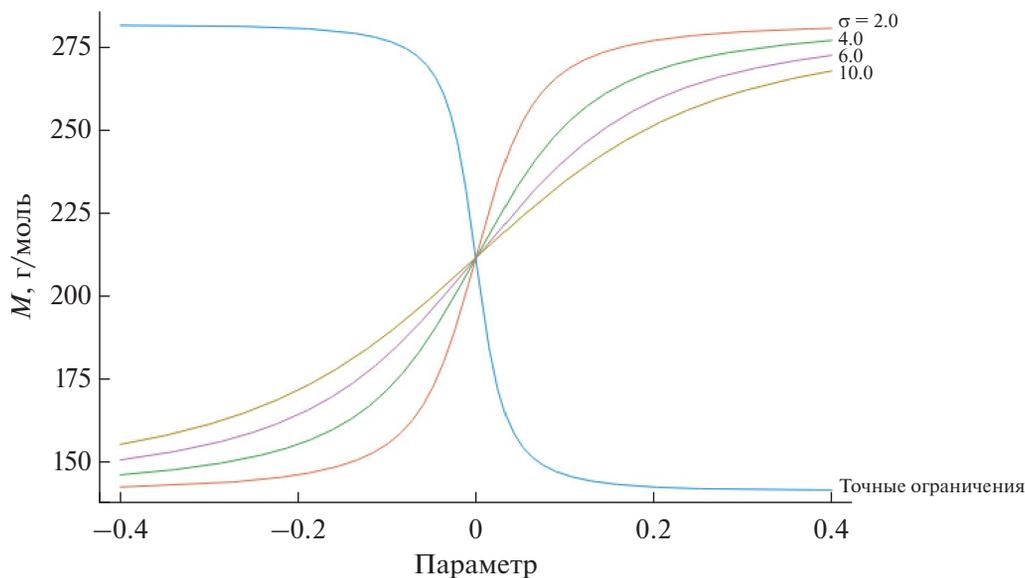


Рис. 4. Зависимость расчетной молекулярной массы смеси в зависимости от значения параметра модели для точных и неточных ограничений.

$$\frac{\partial x_i}{\partial \lambda_1} = \left( \frac{e^{-\lambda_1 M_i - \lambda_2 \omega_i}}{\sum e^{-\lambda_1 M_i - \lambda_2 \omega_i}} \right)'_{\lambda_1} = -M_i x_i + x_i \bar{M} = x_i (\bar{M} - M_i), \quad (34)$$

$$\frac{\partial x_i}{\partial \lambda_2} = \left( \frac{e^{-\lambda_1 M_i - \lambda_2 \omega_i}}{\sum e^{-\lambda_1 M_i - \lambda_2 \omega_i}} \right)'_{\lambda_2} = -\omega_i x_i + x_i \bar{\omega} = x_i (\bar{\omega} - \omega_i). \quad (35)$$

Тогда элементы Якобиана можно найти как

$$\frac{\partial \bar{M}}{\partial \lambda_1} = \left( \sum_{i=1}^N M_i x_i \right)' = -\text{Var}(M), \quad (36)$$

$$\frac{\partial \bar{M}}{\partial \lambda_2} = \left( \sum_{i=1}^N M_i x_i \right)' = \sum_{i=1}^N M_i x_i = \sum_{i=1}^N M_i x_i (\bar{\omega} - \omega_i) = \bar{M} (\bar{\omega} - \omega), \quad (37)$$

$$\frac{\partial \omega}{\partial \lambda} = \left( \frac{\sum M_i \omega_i x_i}{\sum M_i x_i} \right)' = \left( \frac{\overline{M\omega}}{\bar{M}} \right)' = \frac{\overline{M\omega}' \bar{M} - \overline{M\omega} \bar{M}'}{\bar{M}^2} = \frac{\overline{M\omega}' - \omega \bar{M}'}{\bar{M}}, \quad (38)$$

$$\frac{\partial \omega}{\partial \lambda_1} = \frac{\overline{M\omega}' + \omega \text{Var}(M)}{\bar{M}} = \frac{\overline{M\omega}' - \overline{M^2 \omega}}{\bar{M}}, \quad (39)$$

$$\frac{\partial \omega}{\partial \lambda_2} = \frac{\overline{M\omega}' + \omega \text{Var}(M)}{\bar{M}} = \omega \left( \bar{\omega} + \frac{\text{Var}(M) - \overline{M\omega^2}}{\bar{M}} \right). \quad (40)$$

Тогда Якобиан будет равен

$$J = \begin{bmatrix} -\text{Var}(M) & \bar{M}(\bar{\omega} - \omega) \\ \frac{\overline{\omega M^2} - \bar{M}^2 \bar{\omega}}{\bar{M}} & \omega \left( \bar{\omega} + \frac{\text{Var}(M) - \bar{M} \bar{\omega}^2}{\bar{M}} \right) \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}, \quad (41)$$

а ковариационная матрица параметров

$$\text{cov}(\Delta p) = \frac{1}{(\det J)^2} \begin{bmatrix} D & -B \\ -C & A \end{bmatrix} \begin{bmatrix} \sigma_M^2 & 0 \\ 0 & \sigma_\omega^2 \end{bmatrix} \begin{bmatrix} D & -C \\ -B & A \end{bmatrix} = \frac{1}{(\det J)^2} \begin{bmatrix} D^2 \sigma_M^2 + B^2 \sigma_\omega^2 & -(DC \sigma_M^2 + AB \sigma_\omega^2) \\ -(DC \sigma_M^2 + AB \sigma_\omega^2) & C^2 \sigma_M^2 + A^2 \sigma_\omega^2 \end{bmatrix}. \quad (42)$$

А значит,

$$\sigma_{\lambda_1}^2 = \frac{D^2 \sigma_M^2 + B^2 \sigma_\omega^2}{(\det J)^2}, \quad (43)$$

$$\sigma_{\lambda_2}^2 = \frac{C^2 \sigma_M^2 + A^2 \sigma_\omega^2}{(\det J)^2}. \quad (44)$$

Видно, что дисперсия параметров может быть выражена как взвешенная сумма дисперсий измерений.

Может показаться достаточно контринтуитивно, что в выражении для неопределенности в параметре, отвечающем за молекулярную массу, фигурирует произведение неопределенности в экспериментальном определении молекулярной массы и производной массовой доли углерода по параметру, отвечающему за массовую долю углерода, но если переписать это выражение в следующем виде:

$$\sigma_{\lambda_1}^2 = \frac{\sigma_M^2}{\left(\frac{\det J}{D}\right)^2} + \frac{\sigma_\omega^2}{\left(\frac{\det J}{B}\right)^2} = \frac{\sigma_M^2}{\left(A - \frac{BC}{D}\right)^2} + \frac{\sigma_\omega^2}{\left(\frac{AD}{B} - C\right)^2}, \quad (45)$$

то можно заметить, что это достаточно естественное обобщение случая с одной переменной.

Ковариация параметров равна:

$$\text{cov}(\lambda_1, \lambda_2) = \frac{-(DC \sigma_M^2 + AB \sigma_\omega^2)}{(\det J)^2}, \quad (46)$$

и коэффициент корреляции между параметрами будет равен:

$$r(\lambda_1, \lambda_2) = \frac{-(DC \sigma_M^2 + AB \sigma_\omega^2)}{\sqrt{D^2 \sigma_M^2 + B^2 \sigma_\omega^2} \sqrt{C^2 \sigma_M^2 + A^2 \sigma_\omega^2}}. \quad (47)$$

Это выражение равно единице (по модулю) лишь тогда, когда якобиан вырожден, т.е. столб-

цы/строки являются линейно зависимыми. Это легко заметить, поняв, что данное выражение представляет собой скалярное произведение нормированных векторов.

#### Стохастическая реконструкция

Рассмотрим ранее приведенный пример со смесью нормальных парафинов, но в этот раз воспользуемся стохастической реконструкцией.

Предполагая, что длина парафинов подчиняется гамма-распределению, можно добиться любых допустимых значений молекулярной массы и доли углерода в смеси, варьируя среднее и дисперсию (рис. 5).

Можно заметить достаточно общую тенденцию для обоих свойств: они довольно сильно

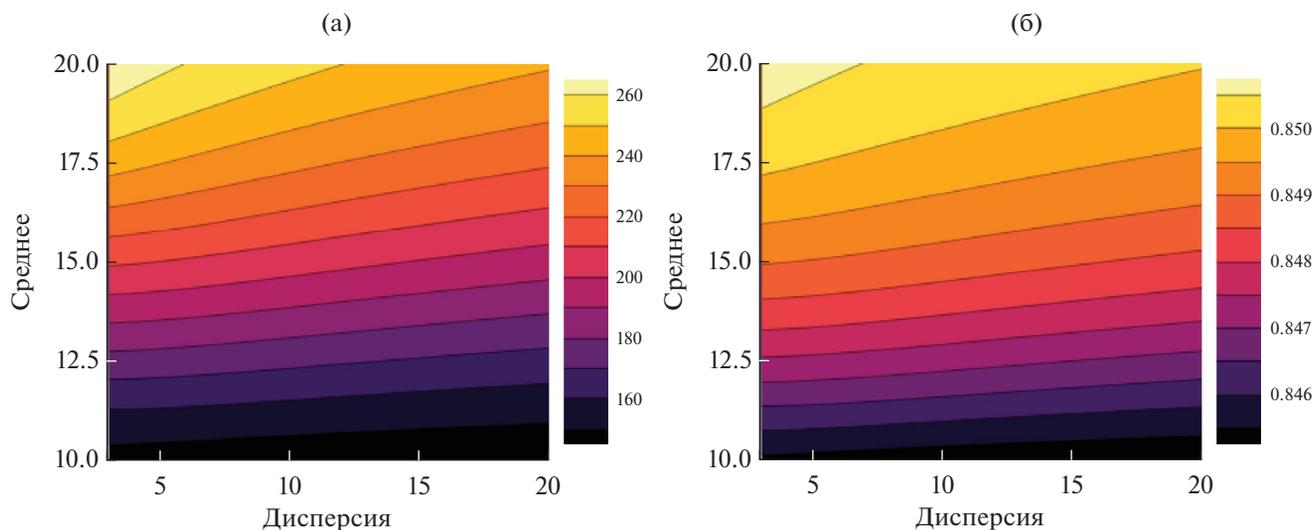


Рис. 5. Средняя молекулярная масса (а) и массовая доля углерода (б) смеси в зависимости от параметров модели.

меняются с изменением среднего и почти не меняются с изменением дисперсии, что не кажется удивительным, поскольку оба свойства в некотором смысле измеряют “среднее” поведение смеси.

В связи с этим сложно ожидать хоть какой-либо точности в определении параметров. Определитель матрицы Якоби в данном случае для большинства значений составляет порядка  $10^{-15}$  (т.е. близок к машинной точности), и коэффициент корреляции равен 1 (самое меньшее значение отстояло от единицы не более чем на  $5 \times 10^{-16}$ ). Расчетные значения логарифмов среднеквадратичных отклонений при  $\sigma_M = 1$  г/моль

и  $\sigma_\omega = 0.1$  мас. % показаны на рис. 6. Можно отчетливо видеть численную неустойчивость, вызванную почти сингулярным якобианом.

Если же использовать другое свойство, которое в большей степени характеризует дисперсию, то ошибка определения параметров будет заметно ниже. Например, если измерить долю отгона при двух температурах, можно найти произведение долей самой легкой и самой тяжелой частей. Несложно заметить, что для того, чтобы такая величина была большой, необходимо значительное количество легкой и тяжелой частей одновременно. Расчеты точности параметров модели в этом случае приведены на рис. 7. Вто-

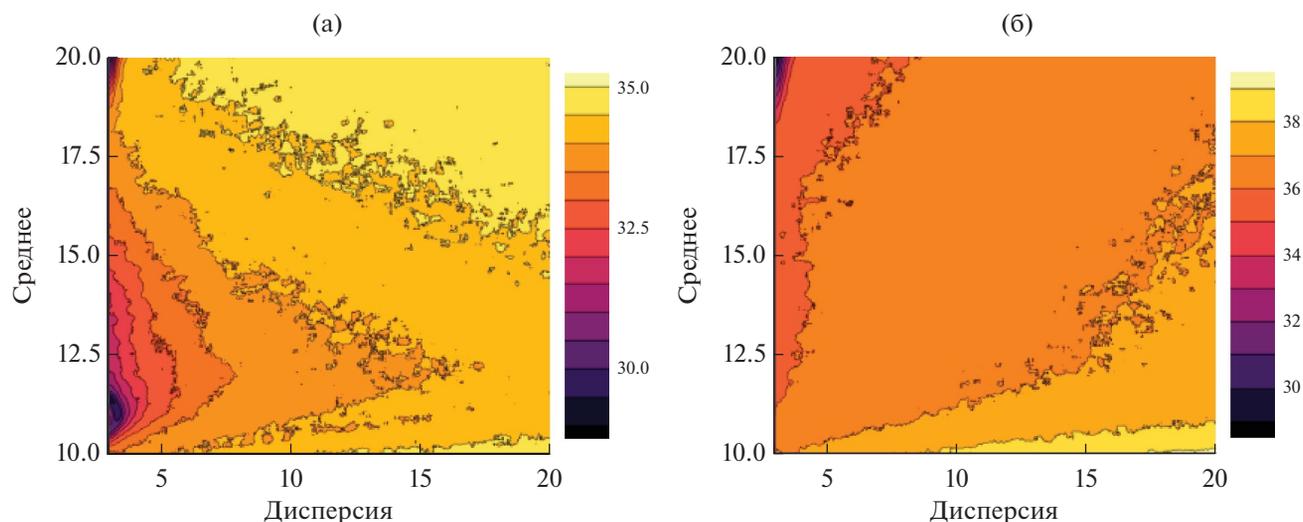


Рис. 6. Зависимость логарифма среднеквадратичного отклонения среднего (а) и дисперсии (б) в зависимости от истинных параметров модели.

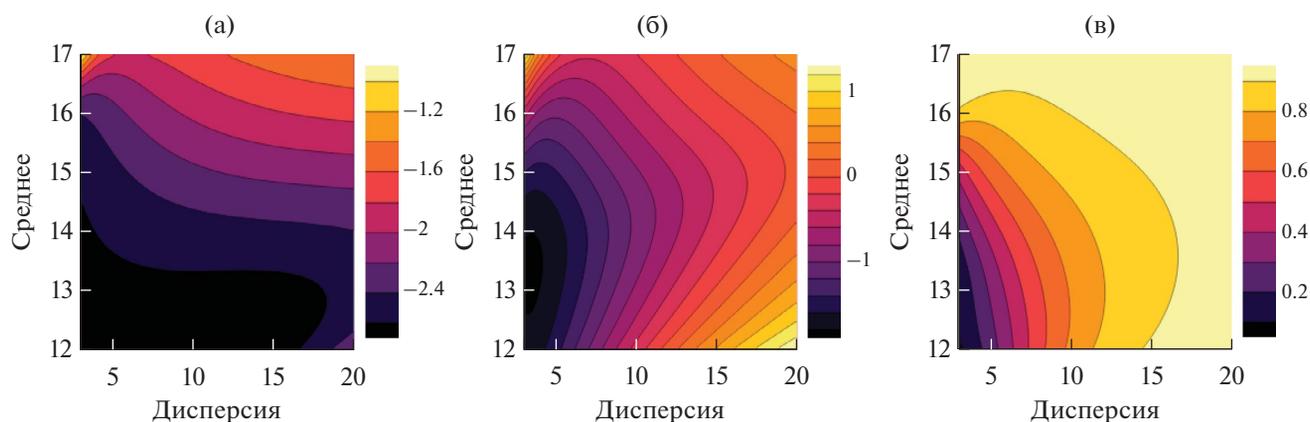


Рис. 7. Зависимость логарифма среднеквадратичного отклонения среднего (а), дисперсии (б) и коэффициент корреляции (в).

рым свойством выступало произведение суммарных концентраций  $C_{10}-C_{12}$  и  $C_{18}-C_{20}$ .

## РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

Сравнивая рис. 6 и 7, можно заметить, что рис. 7 существенной численной неустойчивости не наблюдается. Кроме того, значения оценки погрешностей меньше почти на 16 порядков.

Важно понимать, что оба набора данных могут быть с легкостью использованы для реконструкции и приведут к каким-то значениям параметров, которые не противоречат данным, но при этом в одном случае это решение будет значительно менее локализовано, что усложнит применение реконструированного состава для целей моделирования кинетики.

Предлагаемый метод не только дает возможность установить ситуации, в которых параметры модели на самом деле могут быть определены из экспериментальных данных, но и оценить то, насколько им можно доверять, тем самым позволяя подобрать методы анализа, необходимые для нахождения концентраций с требуемой точностью.

## ЗАКЛЮЧЕНИЕ

Использование реконструированных составов для описания кинетики может очень сильно помочь в построении детальных кинетических моделей и дальнейшей оптимизации процессов, однако необдуманное применение этих подходов способно привести к абсурдным результатам. В связи с этим, необходимо тщательное исследование привлекаемых моделей.

В работе предложен достаточно общий взгляд на методы реконструкции и предложена формула, позволяющая оценивать точность найденных параметров. Несмотря на изобилие математических выкладок в самой работе, применение выдвинутой формулы на практике не должно вызывать трудностей при применении прямого автоматического дифференцирования / (многомерных) дуальных чисел для нахождения Якобиана. Численная нестабильность, возникающая при высокой корреляции выбранных методов анализа, сама по себе может свидетельствовать о малой надежности найденных параметров.

Знание точности параметров позволяет определить, в каких диапазонах могут варьироваться концентрации компонентов, не противореча экспериментальным данным, и соответственно решать, насколько выбранная модель может быть использована для описания кинетики.

## ФИНАНСИРОВАНИЕ

Работа выполнена при поддержке Министерства науки и высшего образования Российской Федерации в рамках государственного задания для Института катализа им. Борескова (проект FWUR-2024-0037).

## КОНФЛИКТ ИНТЕРЕСОВ

Авторы заявляют об отсутствии конфликта интересов, требующего раскрытия в данной статье.

## СПИСОК ЛИТЕРАТУРЫ

1. Зайнуллин Р.З., Коледина К.Ф., Губайдуллин И.М., Ахметов А.Ф., Коледин С.Н. // Кинетика и катализ. 2020. Т. 61. № 4. С. 550. (Zaynullin R.Z., Koledina K.F.,

- Gubaydullin I.M., Ahmetov A.F., Koledin S.N.* Kinetic Model of Catalytic Gasoline Reforming with Consideration for Changes in the Reaction Volume and Thermodynamic Parameters // *Kinet. Catal.* 2020. V. 61. P. 613. <https://doi.org/10.1134/S002315842004014X>)
2. *Заварухин С.Г., Яковлев В.А.* // *Кинетика и катализ.* 2021. Т. 62. № 5 С. 647. (*Zavarukhin S.G., Yakovlev V.A.* Mathematical Modeling of the Nonisothermal Pyrolysis of Sorghum Biomass Based on a Three-Component Kinetic Model // *Kinet. Catal.* 2021. V. 62. P. 688. <https://doi.org/10.1134/S0023158421050128>)
  3. *Zagoruiko A.N., Belyi A.S., Smolikov M.D., Noskov A.S.* // *Catal. Today.* 2014. V. 220–222. P. 168.
  4. *Temkin O.N., Zeigarnik A.V., Kuzmin A.E., Bruk L.G., Slivinskii E.V.* // *Russ. Chem. Bull.* 2002. V. 51. № 1. P. 1.
  5. *Ren Y., Liao Z., Sun J., Jiang B., Wang J., Yang Y., Wu Q.* // *Chem. Eng.* 2019. V. 257. P.761.
  6. *Hudebine D., Verstraete J.J.* // *Oil Gas Sci. Technol.* 2011. V. 66. № 3. P. 437.
  7. *Neurock M., Libanati C., Nigam A., Klein M.T.* // *Chem. Eng. Sci.* 1990. V. 45. № 8. P. 2083.
  8. *Neurock M., Nigam A., Trauth D., Klein M.T.* // *Chem. Eng. Sci.* 1994. V. 49. № 24. P. 4153.

## A General Approach to Molecular Reconstruction Accuracy Estimation

N. A. Glazov<sup>1, \*</sup>, A. N. Zagoruiko<sup>1</sup>

<sup>1</sup>*Boreskov Institute of Catalysis, SB RAS, 5 Lavrentiev Ave.,  
Novosibirsk, 630090 Russia*

*\*e-mail: glazov@catalysis.ru*

A generalization to a few molecular reconstruction methods has been proposed along with a general approach for the accuracy estimation of molecular reconstruction methods. The proposed algorithm estimates the covariance matrix of model parameters based on the matrix of calculated property derivatives with respect to the model parameters and the covariance matrix of measurement error, which allows one to estimate the concentration range in which the solution will not violate collected experimental data. The proposed algorithm is especially easy to use when forward automatic differentiation is employed to find derivatives.

**Keywords:** mathematical modeling, molecular reconstruction, entropy maximization, stochastic reconstruction, accuracy

### FUNDING

The article was supported by program no. FWUR-2024-0037 of the Ministry of Science and Higher Education of Russia for Federal Research Center Boreskov Institute of Catalysis.