

Метод фундаментальных параметров в рентгеновской спектроскопии как задача статистической оценки параметров при наличии ограничений. Критический анализ ситуации.

Eduard Shekhter,

The free researcher, Germany, Linsingenstr. 34, 30163 Hannover,

ed.shekhter@gmail.com

www.chekhter.de

Резюме

С момента своего возникновения в 60 – х годах прошлого века метод описания генерации характеристического рентгеновского спектра с помощью физической модели при облучении вещества возбуждающим излучением достиг таких поразительных успехов, что измерения на образцах сравнения были заменены на расчетные спектры (безэталонный анализ). В противоположность этому, методы решения уравнений количественного анализа остались на эмпирическом уровне. Не определено четко математически что понимается под решением, нет теоретической оценки точности решения в многокомпонентном случае. В статье критически рассматривается сложившаяся ситуация, предлагается определение понятия решения с помощью теории статистической оценки параметров при наличии ограничений. Предлагается решение поставленной задачи в случае многокомпонентного образца. Предлагаемое решение является наилучшим в смысле минимума дисперсии оценки. Кроме того, оценка является несмещенной при осреднении по шумам наблюдений. Вычисляется дисперсионная матрица решения как теоретическая граница Крамера – Рао. Предлагаемый подход открывает возможности применения методов теории статистической оценки параметров в количественной рентгеновской спектроскопии.

" Things should be described as simply as possible, but no simpler" Albert Einstein

1 Введение

В своем рассмотрении я объединяю методы рентгенофлуоресцентной и рентгеновской электронно-зондовой микроскопии без обсуждения существующих различий. Методы различаются глубиной и механизмом генерации характеристического излучения. Общим является описание явлений с помощью теоретической параметрической модели (метод фундаментальных параметров). В случае количественного рентгеновского анализа роль неизвестных оцениваемых параметров играют весовые концентрации компонент образца. Таким образом, математически мы имеем дело с задачей векторного нелинейного анализа.

Так как аналитический сигнал формируется как отношение интенсивностей пиков характеристического излучения к аналогичной интенсивности, измеренной на образце сравнения (или рассчитанной), с целью исключения приборной константы, то уравнения связи можно записать

$$\frac{c_i}{c_i^{st}} = \frac{I_i}{I_i^{st}} f_i(c_1, \dots, c_n, P_1, \dots, P_m) \quad (1)$$

Здесь $\frac{c_i}{c_i^{st}}$ относительные концентрации, $\frac{I_i}{I_i^{st}}$ отношение интенсивности излучения

характеристической линии, исправленной на фон сплошного излучения, к аналогичному излучению от эталонного образца, $f_i(c_1, \dots, c_n, P_1, \dots, P_m)$ - сложная функция, скорее алгоритм, описывающая механизм взаимодействия. Она зависит от n концентраций элементов и m физических параметров. Вид этой функции определяется ситуацией физического эксперимента и предпочтениями исследователей при выборе конкурирующих моделей, дающих примерно одну и ту же точность при расчетах. За подробностями я отсылаю к работе – обзору [1].

При некоторых предположениях о механизме взаимодействия элементов образца, например отсутствии или незначительности поправки на флуоресценцию, формулу (1) можно переписать в виде

$$\frac{c_i}{c_i^{st}} = \frac{I_i}{I_i^{st}} \left(\sum_j \alpha_{ij} c_j \right) \quad (2)$$

Поправка является однородной функцией концентраций. Подобную формулу предложил еще Кастен. В рентгенофлуоресцентном анализе она не нашла дальнейшего применения, а в рентгеновском микроанализе известна как поправка Зиболда – Огилви. Ее аналог, поправка Бенса – Алби применяется широко при расчете геологических образцов.

Если в (2) сгруппировать члены, то уравнения можно переписать в виде однородного матричного уравнения

$$\mathbf{Bc} = \mathbf{0} \quad (3)$$

Здесь \mathbf{B} ($n \times n$) матрица с элементами $b_{ij} = \begin{cases} k_i \alpha_{ii} - 1, & i=j \\ k_i \alpha_{ij}, & i \neq j \end{cases}$, не зависящими от концентраций,

$k_i = \frac{I_i}{I_i^{st}}$ относительные интенсивности характеристического излучения, $\mathbf{c} = [c_1, \dots, c_n]'$ вектор

весовых концентраций. Если детерминант матрицы \mathbf{B} не равен нулю, то уравнение имеет

только нулевое решение, то есть все искомые концентрации равны нулю! Но все сразу не могут равняться нулю, это бессмысленно, ведь должно априори выполняться

$$\sum_i c_i = 1 \quad (4)$$

Сумма всех весовых концентраций должна составлять 100%.

Это породило научную дискуссию на рубеже веков, ситуация обсуждалась неоднократно в литературе, но действительного понимания не было достигнуто. Под решением по-прежнему понимали алгебраическое решение нелинейной системы (1) или (2). При этом понимали его по-разному.

Так, автор Зиболд в статье [3], посвященной точности и разрешающей способности рентгеновского микроанализатора, рассматривал двухкомпонентную систему, описываемую уравнением (2). Он молча предположил, что эта система, состоящая из двух уравнений, вырождена, выбросил второе уравнение из (2), вместо него взял уравнение (4), произвел с его помощью исключение переменной c_2 в первом уравнении «по разности», $c_2 = 1 - c_1$, получил уравнение относительно c_1 и решил его формульно, $c_1 = f(I_1)$. Второго измерения I_2 как бы не существовало. Он пошел дальше и для этого полученного решения вычислил дисперсию σ_c^2 и определил разрешающую способность и предел обнаружения рентгеновского микроанализатора. То, что он потерял ровно половину имеющейся информации, I_2 , никого не смутило.

В рентгенофлуоресцентном анализе пошли другим путем. Используя уравнение (4), из каждого уравнения в (2) исключили по одной составляющей, получили квадратную неоднородную линейную систему уравнений, которую решают обычными методами (Методы Трейла – Лайчанса, Битти – Брисси). Имеется в виду снова алгебраическое решение задачи. Подобный метод решения является эмпирическим, по поводу его обоснования трудно что – либо сказать.

В микроанализе в многокомпонентном случае тоже применяют эмпирическую процедуру итераций, применяя на каждом шаге итераций комбинацию формул (1), (2) и (4), гиперболическая итерация Хейнрича [13].

В итоге предлагают два ответа с последнего и предпоследнего шага итераций, один точно удовлетворяет уравнению (1), другой точно удовлетворяет уравнению (4). Публикуются оба ответа. Вычислить теоретически точность такой процедуры не представляется возможным. Тогда проделали гигантскую работу по анализу большого количества эталонных образцов с помощью этой процедуры и составили гистограмму относительных отклонений результатов анализа от действительного (химического) содержания элементов в образцах. В среднем результаты анализа давали удовлетворительную точность. Что-то вроде средней по больнице температуры.

Все это сформировалось в середине прошлого – начале нынешнего века и позволило одному из отцов – создателей количественного рентгеновского микроанализа Курту Хейнричу опубликовать очерк, озаглавленный «Золотой век микроанализа» (The Golden Age of Microanalysis), [2]. Статья была опубликована в 2001 году. Автор статьи считал, что все

основные проблемы теории и построения аппаратуры уже решены. Статья содержит прекрасный исторический очерк.

Кроме перечисленных методов, были попытки поиска решения методом наименьших квадратов [17]. С помощью уравнения (4) из уравнений системы (1) исключают одну из искоемых концентраций, получают систему n уравнений для $n-1$ переменных. Ее решают обычными методами решения нелинейных переопределенных систем уравнений. Ответ точно удовлетворяет уравнению (4) и приблизительно уравнениям (1). Авторы были недовольны ответом, так как интенсивности сильных пиков, имеющие большую интенсивность счета и, следовательно, большую статистическую погрешность сильно влияли на точность вычисления малых концентраций.

В рентгенофлуоресцентном анализе при поиске решения применялся метод регуляризации по Тихонову. Кроме того, для поиска решения применялись методы случайного поиска (метод проб и ошибок), методы искусственного интеллекта.

Возможно, что существуют еще другие методы, применяемые фирмами в составе программного обеспечения выпускаемых приборов, которые не публикуются фирмами.

Все перечисленные методы, за исключением упомянутого метода наименьших квадратов, не определяют понятие решения. Видимо, имеется в виду алгебраическое решение системы уравнений (1). Применяемые эмпирические методы решения (методы итераций) не позволяют оценить теоретически погрешность решения.

Особое значение придается полученному в процессе итераций ненормированному результату. Не всегда все присутствующие априорно элементы могут быть зарегистрированы или не все могут быть распознаны при качественном анализе.

Это означает, что сумма зарегистрированных элементов не всегда убедительно равняется 100%. Возникало искушение определить эту сумму из физических уравнений (1). Но эти уравнения сами выведены в предположении справедливости уравнения (4)! Можно ли их применять, отказавшись от предположения (4)?

Еще один важный вопрос, решаемый эмпирически, – момент остановки итерации (момент достижения стационарной точки). На него также нет обоснованного ответа.

Так или иначе, все перечисленные проблемы, с которыми сталкиваются при решении уравнений, упираются в вопрос о существовании и единственности решения уравнений. В своих дальнейших рассуждениях по этому поводу я отталкиваюсь от того факта, что уравнение (4) и уравнения (1) образуют переопределенную систему уравнений. Эта переопределенная система будет совместной при условии «бесшумности», беспогрешности ситуации, как если бы постулируемая математическая модель в точности описывала физическую ситуацию и регистрируемые рентгеновские спектры были «бесшумными». Но так не бывает. Поэтому на практике эта система несовместна и единственного алгебраического решения не существует. Решение нужно определить и понимать как-то по другому.

Кроме того, система уравнений (3) хорошо аппроксимирует более сложную систему уравнений (1) и имеет точное решение, равное нулю. Я предполагаю без доказательств, что два примерно одинаковые уравнения, зависящие регулярированным образом (без скачков значений) от

определяемых параметров (концентраций) должны иметь одинаковые решения. Практически гиперболическая итерация Хейнриха использует именно это обстоятельство, заменяя на каждом шаге итераций сложную систему уравнений (1) системой уравнений (3).

Ну не могут решения двух примерно равных друг другу уравнений (1) и (2) отличаться радикально! Следовательно точное безусловное решение системы (1) равно нулю, либо система является вырожденной. Такое решение нас не устраивает.

Поэтому задача поиска решения уравнений (1) формулируется как статистическая задача оценки параметров при наличии ограничений.

2 Оценка концентраций как статистическая задача оценки параметров при наличии ограничений.

2.1 Статистическая постановка задачи

Обозначим относительные интенсивности характеристического излучения по сравнению с эталонным образцом, входящие в формулу (1), $k_i = \frac{I_i}{I_i^{st}}$ и пусть $F_i(c) = \frac{c_i}{c_i^{st} f_i(c_1, \dots, c_n, p_1, \dots, p_m)}$, тогда уравнение (1) с учетом шумов эксперимента можно переписать

$$\mathbf{k} = \mathbf{F}(\mathbf{c}) + \mathbf{n} \quad (5)$$

где все величины, входящие в формулу, считаются векторными для многокомпонентного случая. Вектор \mathbf{n} обозначает вектор статистических погрешностей, вектор шума с известными статистическими свойствами. Вектор \mathbf{k} это случайный вектор экспериментальных данных, $\mathbf{F}(\mathbf{c})$ наша расчетная физическая модель. Эталоны сравнения чистые элементы, более сложный случай приводится к подобному виду. Таким образом, мы имеем задачу оценки параметров сигнала на фоне шума.

2.2 О модели шумов

Статистические свойства шумов в постулируемой модели зависят от ситуации физического эксперимента. Так, если аналитическим сигналом является отношение $k_i = \frac{I_i}{I_i^{st}}$,

где знаменатель I_i^{st} измеряемая на образце сравнения интенсивность, то статистические свойства такого аналитического сигнала описаны в [4] и обусловлены только статистикой счета квантов. Все аппаратные погрешности в таком разовом эксперименте относятся к систематическим погрешностям. Шум считается распределенным по Гауссу с нулевым средним и известной дисперсионной матрицей $\mathbf{D}_k(\mathbf{c})$ полного ранга. Компоненты вектора шумов считаются некоррелированными и матрица шума диагональна. Дисперсия шума $\sigma_{k_i}^2$ вычисляется в различных руководствах, она зависит от среднего значения аналитического сигнала, то есть от истинного значения концентраций, так как интенсивности счета распределены по Пуассону.

В том случае, когда вместо измеряемого I_i^{st} применяются вычисляемые значения, то шум обусловлен только статистикой счета I_i , а погрешность вычисления I_i^{st} относится к систематическим погрешностям. Шум имеет нулевое среднее, некоррелирован, распределен по Пуассону с известной дисперсией. При достаточных интенсивностях счета можно считать его гауссовским.

При проведении повторных экспериментов на образце (в целях калибровки аппаратуры или в других целях), включая изменяющиеся условия измерений (угол отбора излучения, энергия возбуждающего излучения и пр.) можно считать, что модель шумов, кроме погрешности статистики счета, включает погрешности аппаратных измерений. Шум можно считать некоррелированным, с нулевым средним, но с неизвестной дисперсией, которую надо определить в процессе эксперимента.

Если точнее, то погрешность каждого отдельного измерения будет суммой погрешности за счет статистики счета и аппаратурной погрешности.

2.3 Уравнения нормировки и стехиометрии как линейные ограничения

Часто анализируемые образцы в геологии или полупроводниковые составы равно как сплавы металлов представляют собой соединения с частично известными стехиометрическими формулами. Если образец состоит только из соединения с известной химической формулой, например Al_2O_3 , то из курса химии известно, как найти концентрации элементов из линейных стехиометрических соотношений. Если соединение более сложное, например $(Al_2O_3)_x(MnO)_{1-x}$, где x неизвестно, то можно показать, что искомые концентрации связаны линейными соотношениями

$$\begin{cases} \frac{3}{2} \frac{c_{Al}}{M_{Al}} + \frac{c_{Mn}}{M_{Mn}} - \frac{c_O}{M_O} = 0 \\ c_{Al} + c_{Mn} + c_O = 1 \end{cases} \quad (6)$$

Здесь M_{Al}, M_{Mn}, M_O атомные массы элементов.

Если обозначить: $\mathbf{G} = \begin{bmatrix} 3/(2M_{Al}) & 1/M_{Mn} & -1/M_O \\ 1 & 1 & 1 \end{bmatrix}$ неквадратная матрица, а $\mathbf{u} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

вектор, то наши априорные знания можно записать.

$$\mathbf{Gc} = \mathbf{u} \quad (7)$$

Искомое решение должно подчиняться этим ограничениям. Приведенная формула представляет простейший пример, но по этому принципу могут быть записаны как матричные соотношения и более сложные химические формулы. Подробный вывод смотри в [20].

Мы ограничимся в нашем рассмотрении линейными условиями для получения обозримых результатов. Ясно, что таким же образом могут быть рассмотрены и нелинейные априорные ограничения.

2.4 Неравенство Крамера – Рао и оптимальная (наилучшая) оценка концентраций

Если мы регистрируем сигнал на фоне шума, как это постулировано в разделе 2.1, и производим оценку неизвестных параметров (оценка концентраций), а погрешность оценивания характеризуем дисперсией оценки, то параметры этого сигнала могут быть определены с погрешностью, величина которой не может быть сделана меньше определенной границы. Эта граница называется границей Крамера – Рао. Это утверждение касается несмещенной оценки параметров, для которой $E(\hat{c}) = c$. Здесь \hat{c} какая – либо оценка параметра а c его истинное значение. Эта граница не зависит от способа оценивания параметра, а только от количества информации, содержащемся в сигнале.

Вывод и свойства неравенства Крамера – Рао описаны во многих руководствах по статистическому анализу, например в [6, Multivariate Linear Regression Model] или [7].

Здесь приводится без вывода выражение для неравенства Крамера – Рао для оценки векторного параметра при наличии ограничений (7), по сути точно известных «бесшумных» наблюдений с нулевой дисперсией.

$$\mathbf{D}_g \geq \mathbf{D}_{bond} = [(\mathbf{I} - \mathbf{G}^+ \mathbf{G}) \mathbf{W} \mathbf{W} (\mathbf{I} - \mathbf{G}^+ \mathbf{G})]^+ \quad (8)$$

здесь \mathbf{W} определена как $\mathbf{W} = \mathbf{D}_k^{-1/2} \nabla \mathbf{F}(\mathbf{c})$, матрица $\nabla \mathbf{F}$ это матрица градиента с элементами $(\nabla \mathbf{F})_{ij} = \frac{\partial F_i}{\partial c_j}$, вычисляемыми в точке c истинных концентраций, и \mathbf{D}_k дисперсионная матрица шума с элементами $\sigma_{k_i}^2$, \mathbf{D}_g - дисперсионная матрица вектора оценок концентраций, полученных каким – либо способом. Матрица \mathbf{G} определена уравнением (7). Обозначение A^+ означает псевдоинверсию матрицы по Муру - Пенроузу [8].

Неравенство (8) нужно понимать как неотрицательность матрицы $\mathbf{D}_g - \mathbf{D}_{bond}$, то есть для любой несмещенной оценки концентрации c_i ее дисперсия будет больше, чем предел $(\sigma_g^2)_i > (\sigma_{bond}^2)_i$. Доверительный эллипсоид, сконструированный с помощью матрицы \mathbf{D}_g будет охватывать эллипсоид, построенный с помощью матрицы \mathbf{D}_{bond} . Особенностью этой граничной матрицы в нашем случае является ее вырожденность, сингулярность, со степенью вырожденности, равной количеству условий – бесшумных наблюдений. Свойства этой матрицы будут продемонстрированы далее на примере.

Оказывается возможным построить алгоритм оценки параметров, позволяющий достигнуть эту границу – алгоритм Гаусса – Ньютона для обобщенного метода наименьших квадратов. Условия, при которых возможно построение такой оценки для случая нелинейной оценки параметров даются теоремой Гаусса – Маркова [7]. Оптимальные оценки, для которых дисперсия оценки равна границе Крамера – Рао, называются BLUE по начальным буквам английских терминов, которые переводятся как Наилучшие Линейные Несмещенные Оценки. Это оценка с вычисляемыми статистическими свойствами позволяет строить дальнейший содержательный статистический анализ, конструировать совместные доверительные интервалы и тесты для проверки статистических гипотез.

2.5 Оценка концентраций как статистическая задача оценки параметров при наличии ограничений (Restricted Least Squares Estimator)

Пусть вектор \mathbf{k}^* какая-либо реализация случайного процесса \mathbf{k} , определенного формулой (5). Тогда наилучшая (оптимальная) оценка вектора концентраций может быть получена с помощью минимизации взвешенной суммы наименьших квадратов при наличии ограничений [7].

$$\min_{c:\{Gc=u\}} F(c) = \min_{c:\{Gc=u\}} [\mathbf{k}^* - \mathbf{F}(c)]' \mathbf{D}_k^{-1} [\mathbf{k}^* - \mathbf{F}(c)] \quad (9)$$

Обычно задачу на условный минимум решают с помощью метода множителей Лагранжа. С этой целью функцию $\mathbf{F}(c)$ линеаризуют локально около точки \mathbf{c}_0 начального приближения, $\mathbf{F}(c) = \mathbf{F}(\mathbf{c}_0) + \nabla \mathbf{F}(\mathbf{c}_0)(c - \mathbf{c}_0)$ и решают далее линейную задачу по поиску минимума. Это предполагает вычисление сперва безусловного решения и проектирование его на поверхность ограничений. Полученное значение берут за исходное на следующем шаге итераций.

Все это хорошо функционирует, если безусловное решение существует и единственно. В нашем случае мы не можем этого гарантировать. Кроме того, исходная система уравнений может быть не квадратной, но при этом числа уравнений и дополнительных условий должно быть достаточно для получения однозначного решения. Но получить в этом случае безусловное решение затруднительно.

Поэтому мы применим метод фиктивных или латентных переменных, избавляющий нас от необходимости вычислять безусловное решение [8]. Суть его в том, что сначала с помощью искусственного приема и уравнений (7) параметризуются все возможные решения, подчиняющиеся ограничениям, потом среди них ищутся те, которые минимизируют функционал (9). Метод тождественен способу решения физических уравнений «по разности», когда с помощью условий нормировки и стехиометрии исключают сначала ряд переменных из уравнений, а затем решают переопределенную систему физических уравнений для оставшихся переменных. Метод фиктивных переменных избавляет нас от необходимости выбора переменных, подлежащих исключению. За подробностями определений я отсылаю к [8].

Система линейных ограничений может быть формально записана как $\mathbf{G}(c - \mathbf{c}_0) = \mathbf{u} - \mathbf{G}\mathbf{c}_0$, где вектор \mathbf{c}_0 произвольно выбранный вектор начального приближения, и может быть разрешена относительно вектора $c - \mathbf{c}_0$,

$$c - \mathbf{c}_0 = \mathbf{G}^+(\mathbf{u} - \mathbf{G}\mathbf{c}_0) + (\mathbf{I} - \mathbf{G}^+\mathbf{G})\mathbf{z}$$

где \mathbf{z} вектор произвольных переменных. Если выбрать \mathbf{c}_0 так чтобы $\mathbf{G}\mathbf{c}_0 = \mathbf{u}$, то

$$c - \mathbf{c}_0 = (\mathbf{I} - \mathbf{G}^+\mathbf{G})\mathbf{z} \quad (10)$$

Подставляя (10) в (9), получим

$$\begin{aligned}
& \min_{c:\{Gc=u\}} \left[\mathbf{k}^* - \mathbf{F}(\mathbf{c}) \right]' \mathbf{D}^{-1} \left[\mathbf{k}^* - \mathbf{F}(\mathbf{c}) \right] = \\
& = \min_z \left\{ \left[\mathbf{k}^* - \mathbf{F}(\mathbf{c}_0) - \nabla \mathbf{F}(\mathbf{c}_0)(\mathbf{I} - \mathbf{G}^+ \mathbf{G}) \mathbf{z} \right]' \mathbf{D}_k^{-1} \left[\mathbf{k}^* - \mathbf{F}(\mathbf{c}_0) - \nabla \mathbf{F}(\mathbf{c}_0)(\mathbf{I} - \mathbf{G}^+ \mathbf{G}) \mathbf{z} \right] \right\} = \\
& = \min_z \left\{ \left[\mathbf{D}_k^{-1/2} \left(\mathbf{k}^* - \mathbf{F}(\mathbf{c}_0) \right) - \mathbf{D}_k^{-1/2} \nabla \mathbf{F}(\mathbf{c}_0)(\mathbf{I} - \mathbf{G}^+ \mathbf{G}) \mathbf{z} \right]' \left[\mathbf{D}_k^{-1/2} \left(\mathbf{k}^* - \mathbf{F}(\mathbf{c}_0) \right) - \mathbf{D}_k^{-1/2} \nabla \mathbf{F}(\mathbf{c}_0)(\mathbf{I} - \mathbf{G}^+ \mathbf{G}) \mathbf{z} \right] \right\} \\
& = \min_z \left\{ \left[\mathbf{y} - \mathbf{H} \mathbf{z} \right]' \left[\mathbf{y} - \mathbf{H} \mathbf{z} \right] \right\} \tag{11}
\end{aligned}$$

где $\mathbf{y} = \mathbf{D}^{-1/2} (\mathbf{k} - \mathbf{F}(\mathbf{c}_0))$, $\mathbf{H} = \mathbf{D}^{-1/2} \nabla \mathbf{F}(\mathbf{c}_0)(\mathbf{I} - \mathbf{G}^+ \mathbf{G})$.

Это обычный метод наименьших квадратов по переменной \mathbf{z} , который приводит к оценке

$$\hat{\mathbf{z}} = \mathbf{H}^+ \mathbf{y} \tag{12}$$

Выполнив подстановки, получим

$$\begin{aligned}
\hat{\mathbf{c}}_g - \mathbf{c}_0 &= (\mathbf{I} - \mathbf{G}^+ \mathbf{G}) \hat{\mathbf{z}} = (\mathbf{I} - \mathbf{G}^+ \mathbf{G}) \mathbf{H}^+ \mathbf{y} = \mathbf{H}^+ \mathbf{y} = \left[\mathbf{D}^{-1/2} \nabla \mathbf{F}(\mathbf{c}_0)(\mathbf{I} - \mathbf{G}^+ \mathbf{G}) \right]^+ \mathbf{y} = \\
&= \left[\mathbf{D}^{-1/2} \nabla \mathbf{F}(\mathbf{c}_0)(\mathbf{I} - \mathbf{G}^+ \mathbf{G}) \right]^+ \mathbf{D}^{-1/2} (\mathbf{k}^* - \mathbf{F}(\mathbf{c}_0))
\end{aligned} \tag{13}$$

Мы видим, что полученная оценка линейна относительно шума, $(\mathbf{k}^* - \mathbf{F}(\mathbf{c}_0))$, то есть оценка будет в среднем несмещенной, если среднее значение шума равно нулю. Полученное значение $\hat{\mathbf{c}}_g$ можно выбрать за следующее приближение, $\mathbf{c}_0 = \hat{\mathbf{c}}_g$ и повторить вычисления по формуле (13). Теорема Гаусса – Маркова утверждает, что при определенных предпосылках минимум функционала может быть получен за 2 – 3 итерации [7], смотри далее раздел 3.

2.6 Вычисление дисперсионной матрицы

На каждом шаге итерации можно вычислить дисперсионную матрицу оценки, выражение (8), предполагая, что на последнем шаге итераций полученная оценка близка к истинному значению концентрации. В этом случае дисперсионная матрица оценки равна границе Крамера – Рао, вычисленной в точке оценки [7].

$$\mathbf{D}_g = E \left[(\mathbf{c} - \hat{\mathbf{c}}_g)(\mathbf{c} - \hat{\mathbf{c}}_g)' \right] \approx \left[(\mathbf{I} - \mathbf{G}^+ \mathbf{G}) \mathbf{W}' \mathbf{W} (\mathbf{I} - \mathbf{G}^+ \mathbf{G}) \right]_{c=c_n}^+ \tag{14}$$

Матрица $\mathbf{W} = \mathbf{D}_k^{-1/2} \nabla \mathbf{F}(\mathbf{c})$ определена выше.

В том случае, когда речь идет о повторном эксперименте и мы приписываем инструментальные погрешности того или другого рода «шуму» и дисперсионная матрица экспериментальных данных нам неизвестна по той или другой причине, но мы можем считать, что $\mathbf{D}_k = \sigma^2 \mathbf{I}$, где σ^2 неизвестно, то мы получим из (14)

$$\mathbf{D}_g = \sigma^2 \left[(\mathbf{I} - \mathbf{G}^+ \mathbf{G}) (\nabla \mathbf{F}(\mathbf{c}_0))' \nabla \mathbf{F}(\mathbf{c}_0) (\mathbf{I} - \mathbf{G}^+ \mathbf{G}) \right]^+ = \sigma^2 \mathbf{Q}_1^+ \tag{15}$$

и оценку для неизвестной σ^2

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (k_i - F_i(\hat{\mathbf{c}}_g))^2}{m - p} \quad (16)$$

Здесь m число уравнений физической системы, p число дополнительных условий. Мы можем тогда использовать оценку дисперсионной матрицы

$$\hat{\mathbf{D}}_g = \hat{\sigma}^2 \mathbf{Q}_1^+$$

3 Теорема Гаусса – Маркова и итерационный алгоритм

На основании теоремы Гаусса – Маркова для нелинейных оценок параметров мы можем утверждать:

Решение уравнений метода фундаментальных параметров существует в смысле минимума условного функционала (9). Оно совпадает с алгебраическим решением переопределенной системы уравнений (1), (4), если эта система совместна и ее решение существует.

При этом может быть получена несмещенная оценка решения минимальной дисперсии, так называемая оптимальная оценка, достигающая границу Крамера – Рао. Граница Крамера – Рао может быть достигнута не при любом оценивании параметров, но при помощи оптимального алгоритма обобщенного метода наименьших квадратов (алгоритм Гаусса – Ньютона), критерий (9).

При этом, если бы исходные уравнения были линейные и матрица уравнений не сингулярна, то результат мог бы быть получен за один шаг итераций. Для нелинейной проблемы результат может быть получен за 2 – 3 итерации при выполнении следующих условий:

1. Модель должна быть квазилинейной, то есть всюду плавной функцией определяемых параметров. В нашем случае это так. Функция $\mathbf{F}(\mathbf{c})$ плавно зависит от концентраций, как показывают расчеты, но только при их согласованном изменении, то есть при учете ограничений. Аналогичные результаты содержатся в [12].
2. Начальная точка итераций должна находиться в непосредственной близости от искомого решения и, кроме того, находиться на поверхности ограничений, то есть подчиняться ограничениям (7), как это следует из пункта 2.5. Часто применяемое первое приближение $c_{0i} = k_i / \sum_1^n k_i$ может оказаться слишком далеко от искомого решения в случае сильных взаимодействий. Поэтому мной предлагается более близкая стартовая точка для итераций Гаусса – Ньютона. За стартовую точку принимается решение линейной системы уравнений Зиболда -Огилви, полученное по критерию

$$\min_{\mathbf{c} \in \{G\mathbf{c}=\mathbf{u}\}} \|\mathbf{B}\mathbf{c}\|^2 \quad (17)$$

Решение этой задачи может быть получено формульно и равно

$$\hat{\mathbf{c}}_g = (\mathbf{B}'\mathbf{B})^{-1} \mathbf{G}' [\mathbf{G}(\mathbf{B}'\mathbf{B})^{-1} \mathbf{G}']^{-1} \mathbf{u} \quad (18)$$

Решение приблизительно удовлетворяет уравнению (3) и точно подчиняется условиям ограничений (7), (уравнению (4)). За подробностями вычислений отсылаю к [10].

Фактически это улучшенный алгоритм гиперболической итерации, применяемой при поиске решения в методике Хейнрича, только решение получено за одну итерацию формульно. Предлагаемая формула обобщает результат Зиболда, полученный для бинарного случая, на многомерный, а в одномерном случае совпадает с результатом Зиболда. Полученная оценка не является оптимальной с точки зрения статистики, так как является, в принципе, смещенной, она нелинейна по шумам эксперимента, так как экспериментальные данные входят в матрицу \mathbf{B} . Такая оценка не имеет минимальной дисперсии, но является очень хорошим приближением, как показала практика применения уравнений Зиболда -Огилви (Бенса – Алби). Результат получен в предположении несингулярности матрицы \mathbf{B} , более общий случай смотри [19].

3. Первый шаг итераций должен быть в направлении минимума функционала (9), то есть производиться в плоскости ограничений. Я добился этого с помощью подходящей параметризации решения (латентные переменные). Задача свелась к поиску безусловного минимума на поверхности ограничений, что осуществляется с помощью хорошо известных стандартных алгоритмов [7], смотри раздел 2.5. Тем самым избегается эффект «ныряний», колебаний около искомого решения.
4. Оценка должна быть несмещенной. Это означает, что она не содержит систематической ошибки. Это так, если нет больших ошибок в выборе модели, предположениях о составе образца и выборе уравнений ограничений, играющих по сути роль статистических гипотез. Кроме того, алгоритм оценки должен быть линеен по шумам модели и среднее значение шума равно нулю. Короче, матожидание оценки должно быть равно истинному значению параметра. Это выполняется для предлагаемого алгоритма.
5. Так как на каждом шаге итерации вычисляется дисперсионная матрица решения, то это позволяет построить совместные доверительные интервалы, которым должны принадлежать все погрешности оцениваемых концентраций одновременно. Если на определенном шаге итерации разности всех компонент – концентраций для двух последовательных итераций окажутся внутри этих доверительных интервалов, то формально результат, полученный на последнем шаге итераций, можно считать конечной оценкой и итерацию прекратить. Это является критерием остановки итераций.

4 Обсуждение результатов

Как можно заметить, данная статья имеет характер обзора. В ней не приводится формул конкретных физических моделей. Но эти формулы известны и не могут служить предметом публикации. Эти формулы зависят от вида возбуждающего излучения и модели образца, но проблематика является общей. Кроме того, я ссылаюсь на обзоры по рентгеновской спектроскопии [1, 2, 13].

В ней не приводится доказательств приведенных формул и точных формулировок математических теорем. Эти результаты также являются известными, источники приводятся [6, 7, 8, 9].

В статье нет примеров применения предлагаемых критериев получения решения с целью доказательств преимуществ приведенного способа вычислений. Но этого и не требуется.

Построенные оценки параметров наилучшие по определению. Интерес представляет насколько они лучше и в каких случаях. Проведенные расчеты показали преимущества данного способа оценивания [10, 11]. Кроме того, предлагаемые методы поиска решения многократно проверены в прикладной математике. Используемый в некоторых работах по рентгеноспектральному анализу алгоритм регуляризации решения по Тихонову тождественен предлагаемому аппарату неквадратных матриц и преобразованиям псевдоинверсии Мура – Пенроуза или TSVD [14].

Новым содержанием является только четкое определение понятия решения задачи и критерия получения решения, позволяющего оценивать его точность с помощью вычисляемой дисперсионной матрицы решения. Это определение вводит понятие вероятностной трактовки понятия решения и расширяет тем самым спектр задач, допускающих решение.

Сам описываемый алгоритм получения оптимальной оценки параметров представляет лишь принципиальную канву. При его практическом осуществлении приходится решать множество задач, представляющих лишь технические трудности, которые многократно описаны. Требуется аккуратный алгоритм вычисления матрицы гауссиана, критерий малости собственных чисел при вычислении псевдообратных матриц. Известны методы, увеличивающие скорость сходимости алгоритма итерации Гаусса – Ньютона, известны модификации неравенства Крамера – Рао для оценок со смещением. Наибольший простор имеется для формулировки априорных ограничений как вероятностных предположений или вероятностных гипотез для конкретных приложений. Это уже теория распознавания образов. Некоторые трудности представляет создание алгоритма формулировки стехиометрии в виде матричных соотношений в широком спектре природных образцов со сложным фазовым составом. Применяемые для этого сегодня алгоритмы в составе матобеспечения приборов не выдерживают критики.

Полная задача спектрального анализа неизвестного образца предполагает решение задачи качественного анализа, то есть идентификации линий и установление элементного состава, которая не отделима от задачи количественного анализа, так как все наши априорные предположения об интенсивности пиков и фона зависят от определяемых концентраций.

Здесь адекватным было бы применение метода случайного перебора (искусственного интеллекта) для поиска наилучшего соответствия зарегистрированного спектра теоретическому в комбинации с предлагаемым подходом, так вычисляемые теоретические спектр и фон зависят от истинных концентраций.

Стоит рассмотреть еще одну ситуацию, сложившуюся в спектральном анализе. Зачастую путают при постановке задач количественного анализа модельную и регрессионную постановки задач. Чтобы пояснить это, разделим все такие формулировки на две категории. Первая категория это физические зависимости и данный метод фундаментальных параметров. При этом предполагается некий природный механизм, который мы можем удачно описать с помощью математической модели. Параметрам этой модели мы придаем физический смысл. Модель описывает реальность не совсем точно, а приблизительно, справедлива при определенных допущениях и ограничениях на параметры. Мы можем, задавая величину этих параметров, рассчитать с хорошим приближением измеряемые величины, например, спектры излучения. И наоборот, померяв спектры излучения рассчитать состав вещества. Такая задача называется обратной задачей или задачей оценки параметров модели, она возможна с

погрешностью, обусловленной «шумами» модели. Модель является универсальной, не зависит от прибора или лаборатории, параметрам модели мы придаем физический смысл.

Другой вид зависимостей – это регрессионные зависимости. Они не предполагают никакой природной связи между измеряемыми величинами и переменными параметрами, кроме единственного предположения, что при воспроизведении тех же самых условий эксперимента будут получены те же самые результаты. Зависимость измерений от параметров описывается тогда какой-либо произвольно выбранной функцией, чаще всего (но не всегда) линейной или полиномом, коэффициенты определяются из калибровочного эксперимента по совокупности калибровочных измерений. Полученные кривые используются для предсказания значений. Предсказание тем точнее, чем ближе точка предсказания к области точек, через которые была проведена аппроксимирующая кривая. Число параметров (степень полинома и число коэффициентов) произвольно. Примером может служить кривая траектории ракеты с целью предсказания будущих значений положения. Для каждой новой ситуации (траектория ракеты) аппроксимация должна быть проведена по новой. Для таких вычислений оценить точность метода очень проблематично. Все зависит от конкретной ситуации.

Поэтому вызывает удивление применение в рамках метода фундаментальных параметров таких регрессионных зависимостей для определения разрешающей способности спектральных приборов. Делается это с целью подгонки под определения IUPAC и ISO [15, 16, 17]. Логично было бы делать это, оставаясь в рамках параметрической модели, аналогично тому, как это делал Зиболд, с использованием результатов данной работы.

5 Литература

1. Steinbrecher, Stefan. A unified Monte Carlo approach for quantitative standardless x-ray fluorescence and electron probe microanalysis inside the scanning electron microscope. Dissertation, 2004, <http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:bsz:21-opus-13007>.
2. Kurt F.J. Heinrich. The Golden Age of Microanalysis, *Microsc. Microanal.* 7, 108–118, 2001, DOI: 10.1007/s100050010067
3. Ziebold T O 1967 Precision and sensitivity in electron microprobe analysis. *Anal. Chem.* 39 8581
4. Ancey M, Bastenaire F and Tixier R 1978 Microanalysis and scanning electron microscopy. Proc. summer school, St-Martin-d'Herès, France, Sept. 11-16, 1978. Maurice F, Meny L and Tixier R, eds. (Orsay: Les Editions de Physic, 1979, ISBN 2-902731-03-5) 319
5. R B Marinenko and S Leigh "Uncertainties in electron probe microanalysis" 11th European Workshop on Modern Developments and Applications in Microbeam Analysis IOP Publishing IOP Conf. Series: Materials Science and Engineering 7 (2010) 012017 doi:10.1088/1757-899X/7/1/012017
6. Juan M. Rodriguwez-Poo, Computer-aided Introduction to Econometrics, edited by Springer (2002), 350 p., SBN:354044114X
7. G. A. F. Seber, C.J. Wild, Nonlinear Regression, J.W., N.Y., (1989)
8. Albert, A. (1972) Regression and the Moore-Penrose Pseudoinverse; A.P., N.Y.
9. C.R. Rao, Linear Statistical Inference and Its Application, 2nd ed., J.W., N.Y., (1973)
10. E. Shekhter, Industrial Laboratory, 1986, v. 52, Nr. 5, p.p. 445-449

11. E. Shekhter, G. Edelstein, *Industrial Laboratory*, 1991, v. 57, Nr. 5, p.p. 536-545
12. B. I. Kitov, *X-Ray Spectrometry*, 2000; 29: 285–290
13. *Handbook of Spectrometry: 2nd ed.*, edited by R.E. Van Grieken and A.A. Marcowicz, M.D. Inc., (2002)
14. D. Kahaner, C.B. Moller, S. Nash, *Numerical methods and software*, Prentice-Hall, (1989)
15. International Standard, ISO 3534-1, *Statistics-Vocabulary and symbols*, (1993)
16. IUPAC, *Compendium of Chemical Terminology*, 2nd Edition (1997)
17. Michael Mantler, Naoki Kawahara, *The Rigaku Journal*, vol. 21, No. 2, 2004, 17-25
18. Э. Шехтер, Неравенство Крамера –Рао для оценок с ограничениями и потенциальная точность количественного рентгеноспектрального анализа, www.chekhter.de
19. Э. Шехтер, Обобщение формулы Зиболда на многокомпонентный образец, www.chekhter.de
20. Э. Шехтер, Матричная форма стехиометрических соотношений, www.chekhter.de