

УДК 53.087/088

Два метода сглаживания электронных спектров. Сахарук Т.А.//Научное приборостроение. Электронно-ионная оптика. Л.:Наука, 1989, с. 122-129.

Рассматривается проблема сглаживания электронных спектров путем свертки. Анализируется традиционный метод – полиномиальная локальная аппроксимация как частный случай цифровой линейной фильтрации, приводятся рекомендации по его корректному применению. Предлагается метод, основанный на применении принципов цифровой оптимальной фильтрации, позволяющий исключить субъективный фактор, вносимый пользователем при выборе характеристик фильтра за счет использования априорной информации. Приводятся простые алгоритмы расчета фильтров двумя методами. Лит. – 9 назв., ил. – 3.

ДВА МЕТОДА СГЛАЖИВАНИЯ ЭЛЕКТРОННЫХ СПЕКТРОВ

Любой физический эксперимент находится под влиянием большого числа случайных факторов. Это в полной мере относится и к электронной спектроскопии, где результатом эксперимента является энергетический спектр образца – зависимость числа вторичных электронов, испускаемых в единицу времени, от их энергии.

Источником флуктуаций является как сам объект исследования, так и система регистрации – энергоанализатор и усилительный тракт сигнала. Результаты измерений – энергетические спектры могут быть интерпретированы как реализации случайной функции $f_v(x)$, определенной на пространстве выборок $V = \{v\}$.

Многократное снятие спектра одного и того же образца позволяет получить статистическую оценку исследуемого процесса $\bar{f}_N(x)$ для случая равновероятных выборок:

$$\bar{f}_N(x) = \frac{1}{N} \sum_{v=1}^N f_v(x).$$

Оценка $\bar{f}_N(x)$ обладает свойством асимптотической сходимости к ожидаемому значению $\bar{f}(x) = \lim_{N \rightarrow \infty} \bar{f}_N(x)$, называемому в дальнейшем полезной составляющей сигнала. Однако использование этого метода для получения оценки полезной составляющей сигнала не всегда приемлемо из-за возрастания времени эксперимента, недопустимого, например, при исследовании нестабильного образца. В этом случае для устранения нежелательных флуктуаций приходится прибегать к процедуре сглаживания однократно снятого спектра.

Задача ставится следующим образом. Необходимо по экспериментальным данным $f(x)$ найти оценку $\hat{f}(x)$ функции $\bar{f}(x)$, приближающую последнюю наилучшим образом в смысле некоторого критерия.

Всякая процедура сглаживания использует априорно известную информацию об исследуемом процессе, и чем полнее эта информация, тем эффективнее результат сглаживания. В дальнейшем в качестве такой информации используются следующие факты:

1) функция $f(x)$ может быть представлена в виде

$$f(x) = \bar{f}(x) + n(x), \quad (1)$$

где $n(x)$ – аддитивная помеха, являющаяся реализацией стационарного случайного процесса с нулевым математическим ожиданием, не коррелированного с $\bar{f}(x)$;

2) полезная составляющая есть сумма q спектральных линий (пиков) $A_i(x)$, имеющих максимумы в точках x_{oi} :

$$\bar{f}(x) = \sum_{i=1}^q A_i(x - x_{oi});$$

3) все спектральные линии имеют одинаковую форму и для электронной спектроскопии с достаточной точностью описываются функцией Гаусса

$$A_i(x) = A_{oi} \exp(-4 \ln 2 x^2 / D_i^2), \quad (2)$$

где A_{oi} – амплитуда i -й спектральной линии, D_i – ширина i -й спектральной линии на полувысоте.

В настоящее время широкое распространение для решения задачи сглаживания в электронной спектроскопии получил метод полиномиальной локальной аппроксимации (ПЛА) [1-4]. Однако теоретически необоснованное его использование, не учитывающее имеющуюся априорную информацию, не позволяет получать сколько-нибудь достоверные результаты. Тем не менее, остановимся подробнее на этом методе, чтобы

выработать рекомендации по его корректному применению и найти более общее решение поставленной задачи, позволяющее использовать априорную информацию.

Будем рассматривать только случай цифровой обработки сигналов. Пусть функция $f(x)$ задана N своими значениями на дискретной шкале с постоянным шагом и координатами узлов x_k , то есть величинами $f(x_k)$, обозначаемыми в дальнейшем через $f(k)$.

Согласно идее ПЛА функция $\hat{f}(k)$ может быть представлена полиномом степени M в окрестности $k \in [p-m, p+m]$ каждой точки p , называемой центром аппроксимации. Приближение $\hat{f}(k, p)$ имеет вид

$$\hat{f}(k, p) = \psi^T(p-k) \hat{C}(p), \quad (3)$$

где $\psi(k)^T = [\psi_0(k) \dots \psi_m(k)]$ — вектор-строка полиномиальных линейно независимых функций $\psi_i \in \mathcal{R}^{M+1}$ (t — знак транспонирования); $\hat{C} \in \mathcal{R}^{M+1}$ — вектор-столбец оценки коэффициентов разложения.

Вектор $\hat{C}(p)$ вычисляется из условия минимума среднеквадратичной ошибки аппроксимации и в общем случае зависит от координаты центра p . Представление (3) используется для вычисления оценки в центре аппроксимации $p=k$. Исключение: составляют лишь граничные точки, для которых не может быть построена указанная симметричная окрестность.

Введем следующие обозначения: $\hat{F}(p) = [\hat{f}(p-m, p) \dots \hat{f}(p, p) \dots \hat{f}(p+m, p)]^T$ — вектор-столбец оценок размера $2m+1$; $F(p) = [f(p-m) \dots f(p) \dots f(p+m)]^T$ — вектор-столбец результатов измерений размера $2m+1$. Тогда [5] искомая оценка функции $\hat{f}(k)$ в окрестности точки p имеет вид

$$\hat{F}(p) = \Psi^T (\Psi \Psi^T)^{-1} \Psi F(p),$$

где $\Psi = [\psi(-m) \dots \psi(0) \dots \psi(m)]$ — матрица $M \times (2m+1)$.

Целесообразно вместо системы обычно используемых простейших полиномиальных функций $\psi_i(k) = k^i$ применять в качестве базиса аппроксимации систему ортонормированных $\psi_i^0(k)$ на дискретном множестве $k \in [-m, m]$ функций, для которых

$$\sum_{k=-m}^m \psi_i^0(k) \psi_j^0(k) = \begin{cases} 0, & i \neq j; \\ 1, & i = j. \end{cases}$$

Это позволяет значительно повысить эффективность алгоритма по сравнению, например, с предлагаемым в работе [4] за счет исключения процедуры обращения матрицы $\Psi \Psi^T$.

Для ортонормированного базиса

$$\hat{f}(k, p) = \sum_{j=-m}^m \left(\sum_{i=0}^M \psi_i^0(k-p) \cdot \psi_i^0(j) \right) f(p+j). \quad (4)$$

Обозначив

$$h(\ell, j) = \sum_{i=0}^M \psi_i^0(\ell) \cdot \psi_i^0(j), \quad (5)$$

окончательно, для сглаженного значения в точке k , получим

$$\hat{f}(k) = \sum_{j=-m}^m h(\ell, j) \cdot f(k+j), \quad k = \overline{1, N}, \quad (6)$$

$$\ell = \begin{cases} k-m-1 & k \in [1, m], \\ 0 & k \in [m+1, N-m], \\ k+m-N & k \in [N-m+1, N], \end{cases}$$

где ℓ — смещение точки приближения относительно центра интервала аппроксимации.

Проанализируем полученное выражение. В общем случае при фиксированном ℓ оно может быть записано в виде

$$\hat{f}(k) = \sum_{j=-\infty}^{\infty} h(j) \cdot f(k+j). \quad (7)$$

Это уравнение называется дискретным оператором свертки с конечной импульсной характеристикой $h(j)$. Причем, как показано в [6], оператор с импульсной характеристикой вида (5) является линейным оператором ортогонального проектирования P любой функции $f(k)$ из функционального пространства L_2^1 на линейное подпространство Ω_M , заданное базисом $\{\varphi_i(k), i=1, M\}$. Нуль-пространством $\Omega_M^\perp$ этого оператора является ортогональное дополнение Ω_M до L_2 .

Если выполняются следующие условия для функции (4)

$$\begin{aligned} \bar{f}(k) &\in \Omega_M \\ n(k) &\in \Omega_M^\perp, \end{aligned} \quad (8)$$

то оператор (7) осуществляет точное воспроизведение составляющей $f(k)$ и подавление $n(k)$; квадрат нормы ошибки

$$J = \|P(f) - \bar{f}\|^2 \quad (9)$$

при этом равен нулю. Другими словами, точное восстановление возможно только в случае, если полезная составляющая сигнала является линейной комбинацией базисных функций $\varphi_i(k)$, а помеха ортогональна к ним.

Однако в большинстве практических ситуаций условия (8) не выполняются ни для какого M , и, более того, пространства сигналов $\bar{f}(k)$ и $n(k)$ имеют пересечение. В этом случае задача отыскания необходимого оператора должна решаться как задача минимизации функционала ошибки (9) на множестве параметров M и m .

В работе [3] приводится эмпирическая зависимость для вычисления m , обеспечивающая минимум ошибки при фиксированной степени аппроксимирующего полинома $M = 2$ (использование других степеней считается менее эффективным)

$$m = Ent[0,35 D_{min} / T_K - 0,5].$$

где $Ent[\cdot]$ — целая часть числа; T_K — шаг квантования по энергии; D_{min} — минимальная среди всех пиков полуширина.

Для уменьшения ошибки восстановления в работах [3, 7] предлагается использовать многократное применение процедуры сглаживания с различными весовыми функциями. Однако это приводит к значительным затратам машинного времени. Кроме того, выбор пользователем необходимого числа повторений осуществляется эвристически и вносит субъективный фактор в процесс сглаживания.

Как показано в [5], описанный метод ПДА в общем виде не является оптимальным в глобальном смысле, то есть минимально возможное значение ошибки не может быть достигнуто в классе линейных преобразований, получаемых этим методом. Оптимальное решение можно получить только путем минимизации функционала (9) на множестве всех возможных линейных операторов P . Искомый оператор P_{opt} должен быть таким, чтобы вектор ошибки оценивания $e = P_{opt}(f) - \bar{f}$ был ортогонален к сигналу f (принцип ортогональности [6]). Такой результат интуитивно ясен: если имеется зависимость между ошибкой e и сигналом f , то последующей обработкой можно получить лучшую оценку.

Для практической реализации приведенных выше соображений рассмотрим уравнение свертки в базисе комплексных экспонент:

$$\hat{F}(\omega) = H(\omega) \cdot F(\omega),$$

¹⁾ L_2 — бесконечномерное пространство интегрируемых с квадратом функций, т.е. функций, для которых существует интеграл в смысле Лебега $\int_{-\infty}^{\infty} |f(\xi)|^2 d\xi$.

где $\hat{F}(\omega)$, $H(\omega)$, $F(\omega)$ — Фурье-образы $\hat{f}$, $\hat{h}$ и f , соответственно. В теории систем преобразование Фурье $H(\omega)$ принято называть передаточной функцией, а $\hat{F}(\omega)$ и $F(\omega)$ — частотными спектрами соответствующих сигналов. Такое представление позволяет использовать для решения поставленной задачи фундаментальные результаты теории оптимальной фильтрации. С учетом введенных предположений об аддитивности и независимости помехи и при наличии ограничений вида $H(0) = 1$ (условие несмещенности оценки) минимум функционала (9) достигается, если оператор P задан передаточной функцией

$$H_{опт}(\omega) = \frac{(S_{ff}(0) + S_{nn}(0)) S_{ff}(\omega)}{(S_{ff}(\omega) + S_{nn}(\omega)) S_{ff}(0)}. \quad (10)$$

Здесь $S_{ff}(\omega)$, $S_{nn}(\omega)$ — спектральная плотность мощности полезной составляющей сигнала и помехи, соответственно.

По передаточной функции может быть вычислена весовая функция оператора свертки. Тем самым решается задача синтеза алгоритма сглаживания.

Остановимся на вопросах практической реализации предложенного метода.

Вычислим спектральную плотность мощности полезной составляющей сигнала, полагая, что он описывается последовательностью пиков Гауссовой формы. Можно показать, что спектральная плотность мощности такого процесса имеет вид [6]

$$S_{ff}(\omega) = \frac{1}{2 \cdot \Delta x} \bar{S}(\omega), \quad \omega > 0,$$

где $\bar{S}(\omega)$ — спектральная плотность одного пика; Δx — полуширина пика на уровне 0,01 амплитуды.

Величина $S_{ff}(\omega) \Delta \omega$ имеет смысл мощности (среднего квадрата процесса), приходящейся на интервал частот $\Delta \omega$.

В качестве оценки средней спектральной плотности будем использовать квадрат модуля частотного спектра одиночного пика (2) со средней амплитудой $\bar{A}$ и полушириной D_{min} :

$$S_{ff}(\gamma) = A_s^2 \exp(-D_s^2 \gamma^2),$$

где

$$A_s^2 = \frac{\bar{A}^2 \cdot D_{min}}{T_k} \cdot \frac{\pi}{4 \sqrt{\ln 2 \ln 100}},$$

$$D_s^2 = \frac{D_{min}^2 \cdot \pi^2}{S \ln 2 \cdot T_k^2},$$

$\gamma = \frac{T_k}{\pi} \omega$ — относительная частота.

Проблема вычисления спектральной плотности шума может быть решена путем непосредственного измерения на однородных участках сигнала, взятия преобразования Фурье и усреднения по различным выборкам. В случае отсутствия подобной информации, да и вообще в большинстве практических случаев, можно считать, что шумовая компонента является дискретным белым шумом с дисперсией σ^2 .

Подставляя полученные выражения для спектральной плотности шума и сигнала в (10), получим выражение для передаточной функции оптимального фильтра

$$H_{опт}(\gamma) = \frac{1 + Q}{1 + Q \exp(D_s^2 \gamma^2)},$$

где $Q = \frac{S_{nn}(0)}{S_{ff}(0)} = \frac{\sigma^2}{A_s^2}$.

Если имеет место бесконечно большое увеличение отношения сигнал/шум, т.е. шумовая компонента отсутствует (стремится к нулю), то передаточная функция оптимального фильтра стремится к передаточной функции идеального звена.

Наоборот, если отношение сигнал/шум стремится к нулю, то передаточная функция стремится к функции вида $\lim_{Q \rightarrow \infty} H_{\text{опт}}(\gamma) = \exp(-D_s^2 \gamma^2)$.

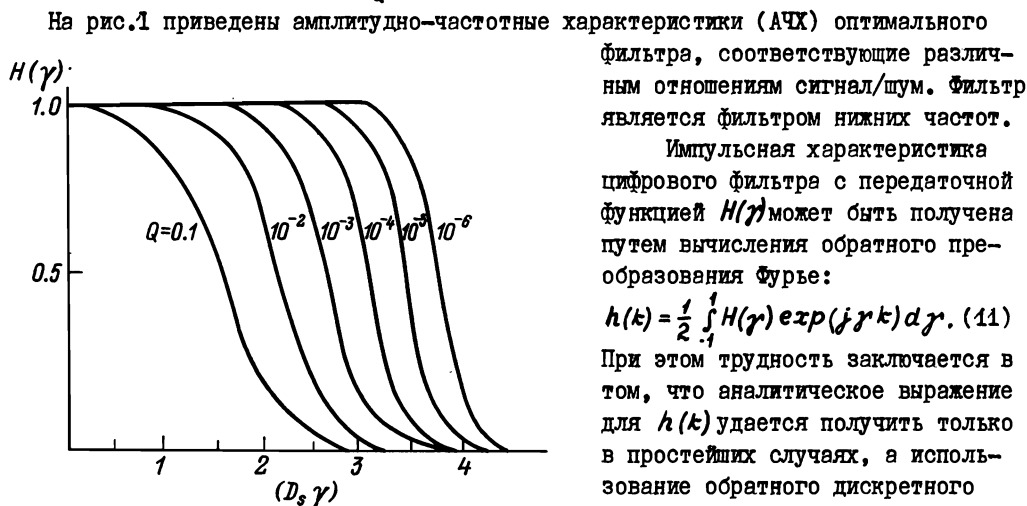


Рис.1. АЧХ оптимального фильтра
 $H(\gamma) = (1+Q)/(1+Q \exp(D_s^2 \gamma^2))$

свойствами. Разработаны эффективные алгоритмы и машинные программы для расчета импульсной характеристики фильтра, передаточная функция которого приближала бы желаемую наилучшим образом в смысле квадратичной нормы, например, второй алгоритм Ремеза, метод частотных выборок и др. [8]. Однако эти алгоритмы сложны и требуют сравнительно больших затрат машинного времени.

В рассматриваемом случае сглаживания электронных спектров передаточная функция оптимального фильтра (10) может быть с успехом аппроксимирована идеальным фильтром нижних частот $H_n(\gamma)$, АЧХ которого равна единице на интервале $(0, \gamma_c)$ и нулю на интервале $(\gamma_c, 1)$.

В соответствии с формулой (11) получим

$$h_n(k) = \int_{-\gamma_c}^{\gamma_c} \exp(j\gamma k) d\gamma = \frac{\sin \gamma_c k}{\gamma_c k}, \quad k > 0.$$

Эта импульсная характеристика бесконечна. Необходимо сделать некоторое усечение полученного ряда Фурье. Однако усечение приводит к появлению пульсаций аппроксимирующей частотной характеристики вблизи точки разрыва и сглаживанию разрыва (рис.2). Ошибка, вносимая усечением ряда Фурье, составляет около 9 %

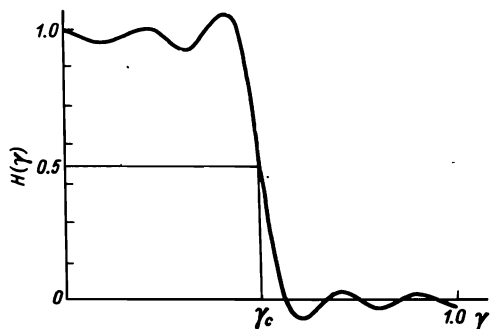


Рис.2. АЧХ фильтра, полученная путем усечения весовой функции идеального фильтра нижних частот

амплитуды в точке разрыва и не зависит от числа членов импульсной характеристики [8, 9]. Эти нежелательные пульсации можно значительно уменьшить, используя импульсную характеристику, полученную методом "окна" [8]. Частотная характеристика такого фильтра имеет трапециевидный вид (рис.3). В работах [8, 9] приведены алгоритмы и рабочие программы расчета весовых функций идеального фильтра; исходными параметрами для расчета являются частота среза γ_c , ширина переходной полосы Δ , ошибка в полосе пропускания δ .

Задача сводится к отысканию частоты γ_c и коэффициента наклона $K = \Delta^{-1}$ фронта трапециoidalной частотной характеристики, аппроксимирующей оптимальную частотную характеристику с наименьшей среднеквадратичной ошибкой. Можно показать, что фронт частотной характеристики, отвечающей поставленному требованию, является касательной к оптимальной частотной характеристике в точке перегиба γ_n .

Точка перегиба определяется из условия равенства нулю второй производной от передаточной функции оптимального фильтра по частоте. Однако получить выражение для γ_n в явном виде не удастся. Решение может быть найдено по методу Зайделя с использованием итерационной процедуры

$$\gamma_n^{(n+1)} = \frac{1}{D_s} \left[\ln \left(\frac{1}{Q} \left(1 + \frac{1}{\gamma_n^{(n)2} - 0,5} \right) \right) \right] \quad (12)$$

и начального приближения

$$\gamma_n = \frac{1}{D_s} \sqrt{\ln \left(\frac{1}{Q} + 2 \right)}. \quad (13)$$

Коэффициент наклона определяется как первая производная от передаточной функции по частоте в точке перегиба

$$K = \frac{2Q(1+Q)\gamma_n D_s \exp(\gamma_n^2 D_s^2)}{1 + Q \exp(\gamma_n^2 D_s^2)} \quad (14)$$

или приближенно

$$K = \frac{D_s \sqrt{\ln(1/Q + 2)(1 + 2Q)}}{2(1+Q)}, \quad (15)$$

тогда для частоты среза получим

$$\gamma_c = \gamma_n + \frac{H_{\text{нгр}}(\gamma_n) - 0,5}{K}. \quad (16)$$

Критерием эффективности применения процедуры сглаживания может служить ожидаемая величина квадрата нормы ошибки восстановления $\hat{\sigma}^2$

$$\hat{\sigma}^2 = 2 \gamma_c \sigma^2 \quad (17)$$

или относительное увеличение отношения сигнал/шум

$$\kappa = \left(\frac{1}{2 \gamma_c} - 1 \right) 100 \%. \quad (18)$$

Малая эффективность имеет место при отсутствии шума (в этом случае сглаживание не нужно), или наоборот, при высоком уровне шума – в этом случае рекомендуется использовать согласованный фильтр

$$H_c(\gamma) = \exp(-D_s^2 \gamma^2 / 2), \quad (19)$$

позволяющий максимизировать отношение сигнал/шум. Импульсная характеристика согласованного фильтра может быть получена с использованием формулы (11). Такая фильтрация решает задачу обнаружения сигнала на фоне белого шума.

Для практического использования частотного метода сглаживания требуется следующая априорная информация: отношение шум/сигнал в исходных данных σ^2/A^2 , число отсчетов, приходящихся на ширину спектральной линии D_{min}/T_K .

Для осуществления сглаживания необходимо выполнить следующие операции.

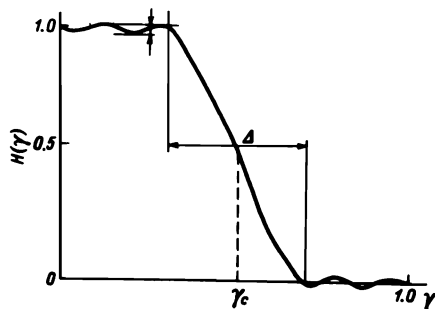


Рис.3. АЧХ фильтра, полученного методом "окна"

1. Вычислить точку перегиба J_n оптимальной АЧХ по формулам (12), (13).
2. Вычислить коэффициент наклона K касательной к оптимальной АЧХ в точке перегиба по формулам (14) или (15).
3. Определить частоту среза γ_c фильтра по формуле (16).
4. Рассчитать эффективность применения процедуры сглаживания по формулам (17), (18).
5. В зависимости от значений величины σ^2/A^2 , δ^2 или κ и конкретных требований к эксперименту принять одно из трех решений:
 - а) отказаться от сглаживания, так как уровень шума достаточно низкий;
 - б) применить согласованный фильтр с целью выявления полезного сигнала на фоне помех (на практике эта ситуация встречается редко и в данной работе подробно не рассматривается. При необходимости рабочие формулы могут быть получены из выражения (19) в соответствии с приведенными выше рекомендациями);
 - в) применить сглаживание оптимальным фильтром.
6. Рассчитать ширину переходной полосы фильтра: $\Delta = K^{-1}$. Задать величину допустимой ошибки в полосе пропускания δ , обычно (0,5 – 3) %. Следует иметь в виду, что уменьшение δ и Δ приводит к увеличению числа членов импульсной характеристики фильтра $2m+1$, от которого линейно зависит время выполнения свертки.
7. Вычислить импульсную характеристику $h(k)$. Исходными параметрами являются γ_c , Δ и δ . Алгоритм расчета и программы на языке ФОРТРАН можно найти в работах [8, 9].

8. Вычислить сглаженную функцию путем свертки

$$\hat{f}(k) = \sum_{j=-m}^m h(j) \cdot f(k+j).$$

Здесь $f(k)$ – исходные данные, заданные на расширенной шкале, т.е. дополненной m узлами слева и справа. Проблема доопределения функции $f(k)$ может быть решена по-разному. Если можно пренебречь фоном, то значения функции в этих узлах полагают равными нулю.

С вычислительной точки зрения эффективность алгоритмов сглаживания определяется временем, затрачиваемым на выполнение операции свертки, поскольку расчет коэффициентов весовых функций как в первом, так и во втором методах занимает незначительную долю машинного времени. Как показывает практика, в этих методах используются примерно одинаковые значения параметра m , а если и имеет место незначительное увеличение во втором методе, то оно окупается качеством и достоверностью сглаживания. Кроме того, алгоритм можно видоизменить так, чтобы предотвратить увеличение значения параметра m свыше установленного предела. Оба метода реализованы на языке ФОРТРАН для ЭВМ СМ1420.

ЛИТЕРАТУРА

1. Анализ поверхности методами оже- и рентгеновской фотоэлектронной спектроскопии/Пер.с англ.Под ред.Д.Бриттса, М.П.Сиха. М.: Мир, 1987.-600 с.
2. Savitsky A., Colay H.J.E.//Anal.Chim.1964. V.36, N 8. P.1627-1639.
3. Proctor A., Sherwood P.M.A.//Anal.Chim.1980. V.52, N 14. P. 2315-2321.
4. Khan A.//Anal.Chem. 1987. V. 59, N 4. P. 654-657.
5. Катковник В.Я. Линейные оценки и стохастические задачи оптимизации (метод параметрических операторов усреднения). М.: Наука, 1976.-488 с.
6. Френкс Л. Теория сигналов/Пер.с англ.М.: Сов.радио, 1974. -344 с.

7. Evans S., Hiorns A.G.//Surface and interface analysis. 1986.V.8.P.71-73.

8. Цифровые фильтры и их применение/Пер.с англ.В.Капшелини, А.Дж.Константи-
тинидис, П.Эмилиани. М.: Энергоатомиздат, 1983.- 360 с.

9. Кайзер, Рид/Приборы для научных исследований. 1977. Т.48, № 11.- С.82-
83.