

В.А. Тартаковский

Условия причинности и демодуляция оптических сигналов с монотонной фазой

Институт оптического мониторинга СО РАН, г. Томск

Поступила в редакцию 12.11.2001 г.

Обсуждается состояние проблемы определения амплитуды и фазы сигнала применительно к анализу полосчатых картин типа интерферограмм. Выделены две задачи, имеющие как теоретическую, так и практическую значимость: условия существования двухполосного спектра у действительного сигнала и оптимальное вычисление преобразования Гильберта. Приводятся конструктивные результаты.

Введение

Сигналы, зависящие от двух переменных и имеющие вид колебания, в том числе замкнутых или незамкнутых полос, наблюдают в оптической интерферометрии. Они могут иметь различное физическое происхождение и формируются при исследовании поверхностей, прозрачных тел, волновых процессов. Подобные сигналы образуются и естественным путем как, например, годовые кольца деревьев, являющиеся индикатором изменений во времени многих природно-климатических факторов. Вариации формы и взаимного расположения полос содержат информацию об изучаемом объекте и могут рассматриваться как пространственная амплитудно-фазовая модуляция.

В общем случае для представления колебательного или волнового процесса в виде двух величин (амплитуды и фазы) требуется непротиворечивое определение этих понятий. Волновое уравнение или уравнение, описывающее колебательный процесс, такого определения не содержат, поэтому необходимы некоторые дополнительные соображения.

Имеется значительное количество работ [2–6, 9, 10, 21, 27, 32, 34, 39], в которых рассматриваются проблемы определения амплитуды и фазы применительно к процессу, зависящему от одной переменной. Различные методы определения полезны в рамках решаемых задач и используемых математических моделей. Однако аналитический сигнал (АС), введенный Д. Габором в 1946 г. [27], получил благодаря работам Д.Е. Вакмана наибольшее распространение.

Картины, представляющие собой полосы, постоянно являются предметом исследований. Их актуальность подтверждается, например, ежегодной конференцией в Германии «Bremer Institut für Angewandte Strahltechnik» («Fringe»). В Томске, в Институте оптики атмосферы, подобные исследования проводились со дня его основания. Вначале С.С. Хмелевцовым, В.В. Покасовым, В.П. Лукиным и О.Н. Емалевым был разработан фазометр и проведены первые

натурные измерения флуктуаций фазы лазерного излучения, распространяющегося в атмосфере.

В 1974 г. автор данной статьи занялся развитием математического описания пространственных интерференционных картин с целью их компьютерного анализа. В дальнейшем, начиная с 1982 г., в Объединенном институте оптики атмосферы СО АН СССР, совместно с Л.А. Пушным и в тесном сотрудничестве с Э.А. Витриченко (сектор академика А.М. Прохорова в Институте космических исследований АН СССР) были разработаны методы высокоточного интерференционного контроля астрономической оптики. Они основывались на формализме аналитического сигнала и дисперсионных соотношений и обеспечивали среднеквадратическую ошибку измерения поверхности меньше сотой доли длины волны [7, 8]. Получивший в настоящее время широкое применение метод демодуляции интерферограмм путем фильтрации в тригонометрическом базисе был предложен этим коллективом в 1982 г. одновременно с авторами работы [38].

С развитием адаптивной оптики впервые потребовалось осуществить фазовую модуляцию световой волны, изменяя в пространстве и во времени форму оптической поверхности. Поэтому возникла необходимость дать непротиворечивое и конструктивное определение фазы световой волны, что было сделано путем обобщения АС на пространственно-временной случай [19].

В настоящее время в Институте оптического мониторинга эти методы продолжают развиваться. Появилась возможность применить их для анализа годовых колец деревьев, которые являются одним из многих источников информации о вековых изменениях температуры, влажности, химического состава почвы и атмосферы. Можно найти дендрооптические аналогии и рассматривать годовые кольца как интерферограмму, образованную некоторыми «экологическим» и «биологическим» полями.

В данном обзоре будут рассмотрены основы формализма аналитического сигнала и связанные с ним дискуссионные проблемы, представлены расширение

возможностей аналитического сигнала и уход от него в частном случае, предложена оптимальная численная реализация базового алгоритма – преобразования Гильберта.

1. Проблемы определения амплитуды и фазы

Простейшая математическая модель сигналов, зависящих от двух переменных, может иметь следующий вид:

$$G(x, y) = B [|1 + W(x, y)|^2] = B [1 + a^2(x, y) + 2a(x, y) \cos\phi(x, y)]. \quad (1)$$

Здесь $W(x, y)$ – комплексное объектное поле с амплитудой $a(x, y) < 1$ и фазой $N(x, y)$; B – оператор, характеризующий нелинейный детектор. Преобразование, которое он реализует, может быть обратимым или необратимым, например бинарным.

Фундаментальным свойством, которым обеспечивается полосчатая структура сигнала, является монотонность фазы $N(x, y)$ в параметрическом сечении плоскости xy . При этом экстремумы сигнала, определенного в сечении, должны совпадать с экстремумами сигнала, определенного на всей плоскости. Этим свойством обеспечивается возможность демодуляции сигнала в отдельных независимых сечениях [19]. В связи с этим в дальнейшем будем использовать только один аргумент x .

Функция косинус, входящая в модель сигнала (1), – периодическая и четная. Поэтому немонотонная фаза может дать ту же самую модуляцию сигнала, что и монотонная. Если x_0 есть точка экстремума фазы, то $\cos N(x) = \cos \tilde{N}(x)$, где

$$\tilde{N}(x) = \begin{cases} 2\pi n + N(x), & x < x_0 \\ 2\pi(n+1) - N(x), & x > x_0 \\ \pi(2n+1), & x = x_0 \end{cases} \Big|_{n \in (-\infty, \infty)} \quad (2)$$

Фаза $\tilde{N}(x)$ может быть разрывной в точке x_0 , но ее производная останется непрерывной и ограниченной по модулю той же самой величиной, что и $N(x)$.

Анализируя различные методы определения амплитуды и фазы, Д.Е. Вакман принял во внимание непрерывность и дифференцируемость амплитуды сигнала, независимость фазы от единиц измерения сигнала, совпадение с интуитивным представлением об амплитуде и фазе гармонических колебаний. Выводы были сформулированы следующим образом.

– Для гармонических колебаний допустимые определения дают ожидаемый результат, постоянную амплитуду и линейную фазу.

– Для узкополосных сигналов результаты различных методов могут не совпадать, но при уменьшении относительной ширины спектральной полосы сигнала это несовпадение также уменьшается.

– Некоторые методы приводят к появлению особенностей в амплитуде и фазе при обращении в нуль производных сигнала.

– Непротиворечивое наиболее общее определение амплитуды и фазы достигается с помощью аналитического сигнала.

Аналитический сигнал $W(x)$ строится в виде комплексной функции

$$W(x) = U(x) + iV(x),$$

$$V(x) = \frac{1}{\pi} \text{v.p.} \int_{-\infty}^{\infty} \frac{U(s)}{x-s} ds = \mathbf{H} U(x). \quad (3)$$

Здесь несобственный интеграл определен в смысле главного значения по Коши (v.p.) в тех случаях, когда $s \rightarrow \pm \infty$ и при $x = s$. Мнимая составляющая аналитического сигнала $V(x)$ является Гильберт-трансформантой его действительной части $U(x)$. Оператор \mathbf{H} обозначает преобразование Гильберта по аргументу x . Показано [3], что он является единственным линейным оператором, для которого справедливо равенство

$$\mathbf{H} \cos(\alpha_c x + \phi_0) = \sin(\alpha_c x + \phi_0), \quad (4)$$

где ϕ_0 и $\alpha_c > 0$ есть неизвестные константы, имеющие смысл начальной фазы и несущей частоты. Стало быть, в рамках правила (3) АС вводится единообразно для всех действительных сигналов.

Однако после введения АС амплитуду и фазу вычисляют по известным формулам, содержащим неопределенность в знаке для амплитуды и в количестве π для фазы, поэтому полученные функции не обязательно единственны:

$$a(x) = \pm \sqrt{U^2(x) + V^2(x)},$$

$$N(x) = \arctg \frac{V(x)}{U(x)} \pm n\pi, \quad n = 0, 1, 2, \dots \quad (5)$$

Самое интересное, что знаки и целое число n могут изменяться вместе с аргументом x [6].

Наиболее важное свойство аналитического сигнала – это причинность или каузальность его преобразования Фурье [12]. Причинность вытекает из свойств преобразования Гильберта, эквивалентного умножению в области частот α на функцию $-i \operatorname{sgn} \alpha$. Иными словами, АС, соответствующий данной действительной функции $U(x)$, получается путем обнуления одной половины ее спектра

$$W(x) = \int_{-\infty}^{\infty} (1 + \operatorname{sgn} \alpha) e^{-i\alpha x} d\alpha \int_{-\infty}^{\infty} U(s) e^{i\alpha s} ds = 2 \int_0^{\infty} e^{-i\alpha x} d\alpha \int_{-\infty}^{\infty} U(s) e^{i\alpha s} ds. \quad (6)$$

Основное качество математической модели, ее пригодность для исследования, во многом определяется свойствами тех функций, из которых она составлена. При выборе класса функций для представления физической величины следует принять во внимание

существование преобразования Гильберта, которым вводится аналитический сигнал. Периодические функции представляют особый интерес, так как именно для таких функций предназначен алгоритм быстрого преобразования Фурье (БПФ), который является базовым при численном анализе сигналов. В этой связи какие-либо экспериментальные данные целесообразно аппроксимировать подходящей периодической функцией для получения оптимальных спектральных оценок [16]. Эта задача рассмотрена в разд. 3.

Периодические функции имеют в своем спектре особенности вида δ -функций. По теореме Винера–Пэли–Шварца [23] такие функции имеют финитный спектр и являются целыми аналитическими функциями экспоненциального типа (ЦФЭТ). Более того, если рассматривать только функции, ограниченные на действительной оси, то такая ЦФЭТ будет еще и функцией класса A , по Левину [11], или класса B [33]. Из ограниченности ЦФЭТ следует абсолютная интегрируемость ее преобразования Фурье или спектра, включая случай, когда она сама не является квадратично интегрируемой. Справедливо и обратное утверждение, а так как преобразование Гильберта от ЦФЭТ не нарушает абсолютной интегрируемости ее спектра, то Гильберт-трансформанта также будет ограниченной функцией. Преобразование Гильберта существует и для функций, не имеющих финитного спектра, но с непрерывной производной, а также и для более широкого класса функций, которые удовлетворяют условию Гёльдера. Однако из-за необходимости дискретного представления сигналов эти классы применяться не будут.

Проблемы определения амплитуды и фазы интенсивно обсуждаются, и точку ставить еще рано [6, 9]. Причина в том, что не существует единообразного подхода к построению математических моделей колебаний, волн, детекторов и систем, так как носители сигналов и устройства имеют разную физическую природу. Например, не является принципиальным положение начала координат относительно сигнала в пространстве, но во времени всегда существует начальный момент, до которого ничего не происходило. Однако АС дает отличную от нуля амплитуду еще до этого момента. Поэтому возникают сомнения в целесообразности такой модели и потребовалось ее обоснование.

Аналитическая функция полностью определяется своими нулями, например в форме разложения Адамара. При слабой модуляции нули колебания $W(x)$ расположены в комплексных полуплоскостях, $|W(x)| \neq 0$. По мере усиления модуляции нули появляются на действительной оси x , что реализует неоднозначность, содержащуюся в (5). Таким образом, можно сказать, что нули являются признаком порога сложности явления [17].

При распространении, с увеличением пройденного пути и с усилением флуктуаций среды световая волна переходит этот порог сложности, в ней образуются вихри. В центре вихря амплитуда равна нулю, а фаза неопределена. Но в некоторой окрестности цен-

тра вихря фаза изменяется монотонно по углу и является минимальной [20]. Помимо этого изменяется плотность вероятности флуктуаций амплитуды волны, а некоторые приближенные методы решения волнового уравнения становятся неприменимыми. Из этого следует вывод, что существует общая проблема существенного возрастания сложности математического описания колебательных и волновых процессов после появления нулей. В этом смысле аналитический сигнал не упрощает ситуацию. Тем не менее применение АС для анализа двумерных пространственных сигналов является полезным [19]. Для определенности сформулируем следующие условия и порядок введения аналитического сигнала.

а) Действительный сигнал формируется так, что он имеет двухполосный спектр Фурье, левая и правая спектральные полосы сигнала не пересекаются между собой.

б) Ассоциированный комплексный сигнал образуется путем обнуления одной из спектральных полос удвоенного действительного сигнала.

Далее перейдем к непосредственному исследованию двух практически важных задач.

2. Монотонность фазы и дисперсионная причинность спектра

Наблюдая интерференционную картину $G(x, y)$, трудно сделать вывод о том, имеют ли сечения $\text{Re}W(x)$ двухполосные пространственные спектры, что необходимо для применения аналитического сигнала. Эти сечения могут быть широкополосным сигналом, у которого несущая частота мало отличается от полуширины спектральной полосы. При этом начало координат области частот может оказаться внутри спектральной полосы и возможность применения аналитического сигнала для определения амплитуды и фазы становится проблематичной.

Экспериментально можно наблюдать полный профиль интерференционных полос в линейных сечениях интерферограммы, что обеспечивается монотонностью фазы в этом сечении. Монотонность можно предполагать и в случае годовых колец деревьев в силу неотрицательности их радиального прироста.

Имея в виду, что сдвиг спектра в области частот эквивалентен добавлению линейной функции к фазе сигнала, свяжем выполнение условия причинности для спектра комплексной функции $W(x)$ с наличием монотонности у ее фазы. Воспользуемся неравенством Бернштейна [13, 22] в виде

$$\max \left| \frac{dW(x)}{dx} \right| \leq \alpha_s \max |W(x)|, \quad (7)$$

где функция $W(x) = a(x) \exp iN(x)$ принадлежит к классу функций с финитным спектром; α_s – модуль верхней частоты в спектре этой функции.

В важном частном случае, характерном для интерферограммы, когда $N(x)$ отлична от линейной функции, а амплитуда $a(x)$ постоянна, найдем

$$\max |N'(x)| \leq \alpha_s. \quad (8)$$

По теореме о сдвиге спектра для функции $a(x) \exp i [N(x) + \alpha_s x]$ выполняется условие причинности в частотной области, а из неравенства Бернштейна следует, что фаза $N(x) + \alpha_s x$ заведомо монотонна. Таким образом, для наличия монотонности у фазы сигнала с постоянной амплитудой достаточно, чтобы спектр сигнала был причинным. Очевидно, возможны случаи, когда монотонность присутствует, а причинности нет.

Определим эффективную ширину спектра α_e как дисперсию через нормированный момент второго порядка. Такой подход позволяет рассмотреть в том числе функции, не имеющие финитного спектра. Пусть $W(x)$ – ограниченная сверху и T -периодическая функция. Тогда для нее существует дискретный спектр Фурье S_k . Имея в виду равенство Парсеваля, найдем

$$\begin{aligned} \alpha_e^2 &= \frac{\sum_{k=-\infty}^{\infty} k^2 |S_k|^2}{\sum_{k=-\infty}^{\infty} |S_k|^2} = \frac{\int_0^T |W'(x)|^2 dx}{\int_0^T |W(x)|^2 dx} = \\ &= \frac{\int_0^T [a'^2 + a^2 N'^2] dx}{\int_0^T a^2 dx} < A^2 + \max N'^2(x), \quad (9) \end{aligned}$$

где

$$A^2 = \int_0^T a'^2(x) dx / \int_0^T a^2(x) dx.$$

По аналогии с условием причинности введем условие дисперсионной причинности. Для его выполнения достаточно сдвинуть спектр более чем на свою эффективную ширину α_e в любую сторону, например так:

$$a(x) \exp i [N(x) + x \sqrt{A^2 + \max N'^2(x)}]. \quad (10)$$

При достаточно большом сдвиге монотонность фазы и дисперсионная причинность спектра наблюдаются совместно, как это вытекает из неравенства

$$\alpha_e < \sqrt{A^2 + \max N'^2(x)} \geq \max |N'(x)|. \quad (11)$$

Когда $a = \text{const}$, то $A = 0$ и из (11) получим

$$\max |\phi'(x)| > \alpha_e. \quad (12)$$

Таким образом, чтобы выполнялось условие дисперсионной причинности для спектра сигнала с постоян-

ной амплитудой, достаточно наличия свойства монотонности у фазы сигнала.

Сопоставив (8) и (12), получим

$$\alpha_e < \max |\phi'(x)| \leq \alpha_s, \quad (13)$$

откуда следует, что условие причинности более сильное, чем условие дисперсионной причинности, которое вследствие этого может иметь более широкое применение для построения комплексного сигнала по правилам (см. пп. «а» и «б» разд. 1).

Для характеристики расположения спектра сигнала относительно начала координат полезно ввести численную величину, степень причинности

$$p_c = \pm \sqrt{\left(\sum_{k=1}^{NN} |S_k|^2 - \sum_{k=NN+1}^N |S_k|^2 \right) / \sum_{k=1}^N |S_k|^2}, \quad (14)$$

где $NN = N/2 + 1$ есть частота Найквиста; S_k – дискретный спектр сигнала, вычисленный по алгоритму БПФ. Очевидно, что $|p_c| \leq 1$, а равенство достигается, когда спектр комплексного сигнала удовлетворяет условию причинности.

Наличие монотонности у фазы дает возможность повысить степень причинности у спектра. Пусть фаза монотонна и пусть функция $W(x) = a(x) \exp iN(x)$ не является аналитическим сигналом. Если преобразовать переменную x так, что

$$N(x) = \alpha_c \tau, \quad x = N^{-1}(\alpha_c \tau), \quad (15)$$

то спектр функции $W(\tau)$ станет значительно уже и будет локализован в области вокруг точки $\alpha = \alpha_c$. Ширина этого спектра относительно α_c без учета погрешностей преобразования будет определяться только амплитудой $a(\tau)$, которая мало изменится по сравнению с $a(x)$. При $a(x) = \text{const}$ спектр функции $W(x)$ будет удовлетворять условию дисперсионной причинности, а преобразованная функция $W(\tau)$ будет комплексным гармоническим колебанием с частотой α_c .

Однозначность обратной функции N^{-1} обеспечивается монотонностью прямой функции N . Если к тому же производная $N(x)$ не равна нулю, то обратная функция не будет иметь разрывов, что особенно важно для ее численной реализации.

Такое преобразование \mathbf{E} сжимает периоды колебания, которые больше некоторого среднего периода, и растягивает те из них, которые меньше, а обратное преобразование \mathbf{E}^{-1} возвращает колебание в исходное состояние в соответствии с выражениями:

$$\begin{aligned} \mathbf{E} a(x) \cos N(x) &= a\{N^{-1}(\alpha_c \tau)\} \cos\{N[N^{-1}(\alpha_c \tau)]\} = \\ &= a(\tau) \cos \alpha_c \tau, \\ \mathbf{H} a(\tau) \cos \alpha_c \tau &= a(\tau) \mathbf{H} \cos \alpha_c \tau = a(\tau) \sin \alpha_c \tau, \quad (16) \\ \mathbf{E}^{-1} a(\tau) \sin \alpha_c \tau &= a\{N(x)/\alpha_c\} \sin\{\alpha_c N(x)/\alpha_c\} = \\ &= a(x) \sin N(x). \end{aligned}$$

Здесь второе выражение справедливо при условии, что амплитуда $a(x)$ не содержит частот выше чем α_c , поэтому эта функция вынесена [24] за знак преобра-

зования Гильберта по переменной τ . Рис. 1–3 иллюстрируют обсуждаемые действия.

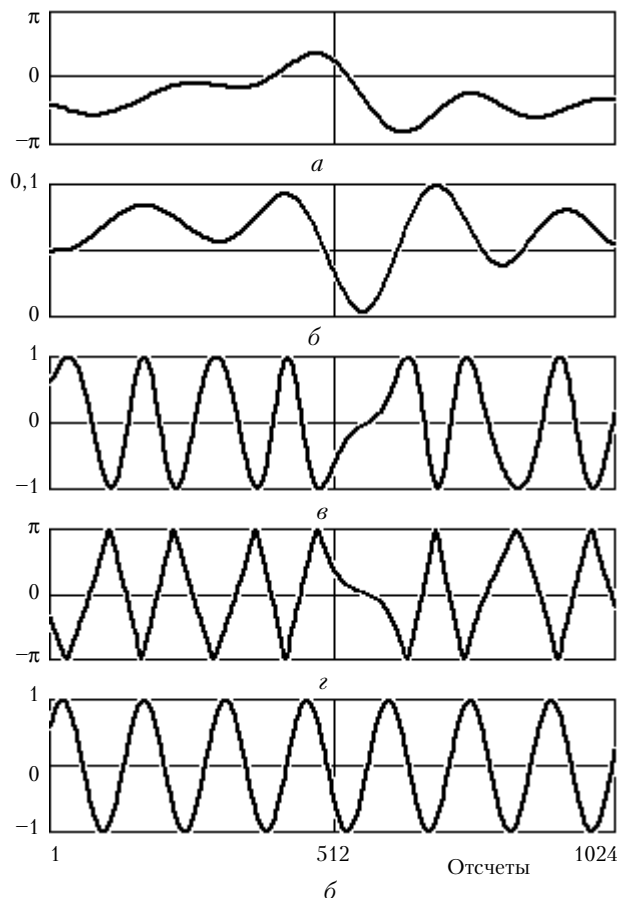


Рис. 1. Преобразования сигнала с монотонной фазой: *a* – монотонная фаза $\phi(x)$ без линейной составляющей; *b* – производная монотонной фазы $\phi(x)$; *v* – сигнал $\cos \phi(x)$; *z* – $\arccos\{\cos \phi(x)\}$; *d* – гармоническое колебание, полученное из сигнала $\cos \phi(x)$ путем операции «сжатие-растяжение»

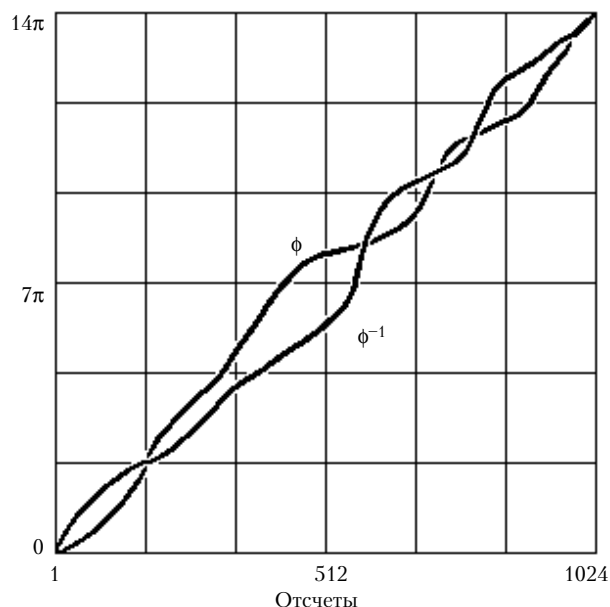


Рис. 2. Монотонная фаза ϕ и обратная ей ϕ^{-1} для сигнала на рис. 1, *v*

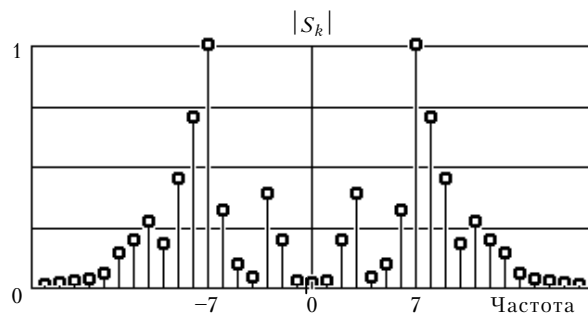


Рис. 3. Модуль преобразования Фурье-сигнала $\cos \phi(x)$ на рис. 1, *v*. Мода на частотах, равных числу периодов сигнала. В окрестности начала координат амплитуда спектра мала. По оси ординат масштаб логарифмический

В общем случае аналогично (3) введем ассоциированный комплексный сигнал $W(x)$ для действительного сигнала $U(x)$ с монотонной фазой по правилу

$$W(x) = U(x) + iV(x), \quad V(x) = \mathbf{E}^{-1} \mathbf{H} \mathbf{E} U(x). \quad (17)$$

В соответствии со способом определения ширины спектра (9) данный комплексный сигнал $W(x)$ будем называть эффективным сигналом. Его следует применять вместо АС при априорно монотонной фазе и при отсутствии двухполосного спектра у функции $U(x)$.

Возникает вопрос о реализации описанных операций. Действительно, для того чтобы определить фазу $N(x)$, имея только $U(x)$, эту фазу уже необходимо знать для осуществления преобразований (15)–(17). Но здесь можно предположить, что величины

$$\phi_0(x) = \arctg \{ \mathbf{H} U(x) / U(x) \} \quad (18)$$

будет достаточно для осуществления начального сжатия спектра. Затем итерационный процесс осуществляется в соответствии с уравнением

$$\begin{aligned} \phi_{n+1}(x) &= \\ &= \arctg \{ \mathbf{H} U[\phi_n^{-1}(\alpha_c \tau)] / U[\phi_n^{-1}(\alpha_c \tau)] \} \Big|_{\tau=\phi_n(x)/\alpha_c}. \end{aligned} \quad (19)$$

Численные эксперименты [19] показали, что четыре итерации по (19) увеличивают точность оценки фазы более чем в десять раз по сравнению с (18).

Монотонная фаза при единичной амплитуде сигнала может быть найдена* непосредственно по действительной части $U(x)$:

$$N(x) = \int \left| \frac{d \arccos U(x)}{dx} \right| dx = \int \frac{|U'(x)|}{\sqrt{1-U^2(x)}} dx. \quad (20)$$

На рис. 1 показана численная реализация этого выражения.

Интересно, что выражение (20) применимо также и для бинарных сигналов. Пусть, например, сигнал $\cos N(x)$ с монотонной фазой подвергается нелинейному преобразованию

$$\mathbf{B} \cos N(x) = \begin{cases} 1, & \cos \phi(x) \geq 0, \\ -1, & \cos \phi(x) < 0. \end{cases} \quad (21)$$

* Результат получен совместно с Ю.Н. Исаевым.

Тогда в соответствии с (20) получим

$$\left| \frac{d}{dx} [\arccos \mathbf{B} \cos N(x)] \right| = \pi \sum_{n=0}^N \delta[\phi(x) - \pi(2n+1)/2] \rightarrow$$

$$\int \rightarrow \pi \sum_{n=0}^N Y[\phi(x) - \pi(2n+1)/2] = N_Y(x), \quad (22)$$

где N – число периодов сигнала.

Таким образом, восстановленная фаза $\phi_Y(x)$ представляет собой сумму функций Хэвисайда, ступенчатую функцию, которая совпадает с исходной фазой $\phi(x)$ в точках, где исходная кратна нечетному числу $\pi/2$, а между ними восстановленная фаза постоянна:

$$N_Y(x) = \left\{ \begin{array}{l} N(x_n), \quad n : N(x_n) = \pi(2n+1)/2 \\ \pi(2n+1)/2, \quad x_n < x < x_{n+1} \end{array} \right\}_{n=0,1,2,\dots} \quad (23)$$

Численная реализация операций (20)–(22) показала отсутствие принципиальных трудностей.

3. Оптимальное вычисление преобразования Гильберта

Алгоритм вычисления преобразования Гильберта, основанный на прямом вычислении дискретной свертки, описан в [30]. Прямое преобразование выглядит следующим образом:

$$V(k\tau) = \left\{ \begin{array}{l} + \frac{2}{\pi} \sum_n \frac{U(n\tau)}{k-n}, \quad k \text{ четное, } n \text{ нечетное;} \\ + \frac{2}{\pi} \sum_n \frac{U(n\tau)}{k-n}, \quad k \text{ нечетное, } n \text{ четное,} \end{array} \right. \quad (24)$$

а обратное имеет симметричную форму

$$U(n\tau) = \left\{ \begin{array}{l} - \frac{2}{\pi} \sum_k \frac{V(k\tau)}{n-k}, \quad n \text{ четное, } k \text{ нечетное,} \\ - \frac{2}{\pi} \sum_k \frac{V(k\tau)}{n-k}, \quad n \text{ нечетное, } k \text{ четное,} \end{array} \right. \quad (25)$$

где τ – полупериод наибольшей частоты в спектре преобразуемой функции; индексы k и n изменяются в пределах интервала определения функций.

По данному алгоритму дискретное представление исходного непрерывного сигнала преобразуется в дискретное представление его Гильберт-трансформанты, которая близка к своему непрерывному прототипу, если:

- отсчеты исходного сигнала получены в соответствии с теоремой Котельникова;
- исходный сигнал хорошо определен своими отсчетами на конечном носителе в том смысле, что вне носителя он значимо не отличается от нуля по какой-либо норме.

Реализация формул (24), (25) в виде программы на Фортране [15] была признана ГосФАП новой в 1978 г. Численные эксперименты показали, что программа вполне работоспособна при выполнении указан-

ных двух условий. Однако нельзя рассчитывать, что точность преобразования может быть высокой в общем случае, так как преобразование Гильберта определено на бесконечном интервале и физически нереализуемо.

Другая возможность создания численного преобразования Гильберта вытекает из его свойств в области частот (6) и сводится к умножению спектра сигнала на знаковую функцию. В одной из ранних работ [25] используется приближенный гармонический анализ для реализации такого подхода.

На данный момент наиболее эффективным и технически обеспеченным алгоритмом для Фурье-трансформации последовательности отсчетов является БПФ [14, 26, 36]. На его основе могли бы быть легко реализованы как преобразование Гильберта, так и другие виды линейных фильтров. Но алгоритм БПФ предполагает, что отсчеты заданы на окружности, стало быть, они могут соответствовать только периодической функции.

Таким образом, существует несоответствие между бесконечной областью определения функций тригонометрического базиса, оператора преобразования Гильберта и конечной областью определения сигнала. Эта область двумерна и, возможно, многосвязная, если речь идет о демодуляции интерферограмм [16, 18].

Описанное несоответствие проявляется в форме краевых выбросов у преобразованной функции и, конечно, у ее фазы. Распространенным методом их подавления в случае спектральных оценок является применение сглаживающих окон. Есть примеры использования окон и в интерферометрии. В работе [31] при применении окна Хэннинга [28] краевой эффект уменьшается приблизительно на три порядка. Но в том случае объектная фаза была постоянна, хотя значения самого сигнала не совпадали на краях интервала, так как число интерференционных полос не было целым. Кроме того, в указанном примере отсутствовал шум. В общем случае окна неэффективны для подавления краевого эффекта у восстановленной фазы. Окна существенно уменьшают амплитуду сигнала на краях интервала определения, и после фильтрации на этих участках происходит относительное увеличение уровня шума. Этот факт проявился в [38], где интерферограмма умножалась на окно Хэннинга. Там краевой эффект заметен на 20% длины интервала определения, а максимальное значение выбросов фазы больше чем $0,2\pi$ рад.

Адекватным методом решения обсуждаемой задачи является оптимальное периодическое продолжение сигнала за пределы области определения с последующим применением БПФ.

В результате численных экспериментов было установлено [8, 19], что монотонности фазы либо возможности привести ее к таковой за счет четности и периодичности косинуса (2) достаточно для введения ассоциированного эффективного сигнала. Это создает широкие возможности для выбора прямолинейных и криволинейных сечений в области определения интерферограммы. Если выбранная кривая сканирования замкнута и принадлежит области определения, то проблема продолжения не возникает, так как сечение является периодической функцией параметра сканиро-

вания. Однако обычно сечения прерываются на внешних границах области определения. Кроме того, при многосвязности появятся разрывы сечения на внутренних границах. В итоге оно может состоять из нескольких конечных фрагментов.

Перейдем к рассмотрению алгоритмов превращения таких фрагментов в периодическую функцию, для которой возможен методически точный алгоритм преобразования Гильберта. Для вывода критерия оптимальности проведем следующие рассуждения. Пусть $U(x)$ есть действительная ограниченная функция, определенная для всех $x \in (-\infty, \infty)$, а сигнал состоит из известных фрагментов этой функции. Пусть далее продолженная функция $\tilde{U}(x)$ совпадает с $U(x)$ в пределах сигнала. Введем функцию $\Omega(x)$, интегрируемую в квадрате и имеющую низкочастотный спектр, не пересекающийся со спектрами функций $\tilde{U}(x)$ и $U(x)$. Примем во внимание свойство [24], позволяющее вынести низкочастотное колебание за знак преобразования Гильберта

$$\mathbf{H}\Omega(x) U(x) = \Omega(x) \mathbf{H} U(x) = \Omega(x) V(x) \quad (26)$$

и равенство энергий $U(x)$ и $V(x)$. Тогда справедливо равенство

$$\begin{aligned} & \int_{-\infty}^{\infty} \Omega^2(x) [U(x) - \tilde{U}(x)]^2 dx = \\ & = \int_{-\infty}^{\infty} \Omega^2(x) \{\mathbf{H}[U(x) - \tilde{U}(x)]\}^2 dx. \end{aligned} \quad (27)$$

Устремим функцию $\Omega(x)$ к единице в пределах известных фрагментов сигнала и к нулю вне этих фрагментов. Левая часть выражения (27) будет при этом стремиться к нулю по определению, а спектр функции $\Omega(x)$ расширяться. Когда спектр расширится настолько, что пересечется со спектрами $\tilde{U}(x)$ и $U(x)$, равенство (27) перестанет быть справедливым. Чтобы нарушение равенства проявилось при возможно меньшем значении его левой части, необходимо обеспечить как можно большую узкополосность функций $\tilde{U}(x)$ и $U(x)$.

Апостериорно можно лишь уменьшить ширину спектра функции $\tilde{U}(x)$ путем следующих операций продолжения:

- более гладкого интерполирования отдельных фрагментов сигнала внутри области определения,
- более гладкого экстраполирования за пределы области определения,
- выбора функциональной зависимости, обеспечивающей эту необходимую гладкость.

Таким образом, для того чтобы Гильберт-трансформанта продолженной функции $\tilde{U}(x)$ была наиболее близка к истинной $U(x)$, необходимо, чтобы операция продолжения обеспечивала минимальную ширину спектральной полосы функции $\tilde{U}(x)$ при за-

данных фрагментах $U(x)$. В этом состоит критерий оптимальности.

В качестве структурных ограничений следует принять во внимание свойства базового алгоритма БПФ [26, 36]. Эта операция определена на окружности. По этой причине сечение следует продолжать на всю бесконечную ось периодически. Кроме того, надо иметь в виду, что алгоритм БПФ перерабатывает массивы вполне определенной длины. Это связано с факторизацией числа, равного длине массива, и с вычислением Фурье-преобразования каждого элементарного массива, длина которого равна соответствующему сомножителю. Обычно этот сомножитель равен двум, а длина всего массива есть целая степень двух. Но всякое составное число может быть единственным образом представлено в виде произведений простых чисел, с точностью до их расположения. Простой сомножитель не может быть большим, так как это уменьшит быстродействие. В работе [36] описан алгоритм и приводится программа на Фортране, использующая для факторизации первые десять простых чисел: 1, 2, 3, 5, 7, 11, 13, 17, 19, 23. Множество составных чисел, образованных из этих сомножителей, обозначим символом P_{10} .

Априорная информация состоит в наличии фундаментального свойства монотонности. Фаза $\phi(x, y)$ в сечении имеет значительную линейную, квадратичную или близкую к ним составляющую, чем и вызвана полосчатая структура интерферограммы. Линейное продолжение этой составляющей с сохранением непрерывности соответствует введенному критерию оптимальности, так как эффективная ширина спектра S_k определяется производной фазы сигнала (9). Следовательно, необходимо создать условия для появления при продолжении дополнительных интерференционных полос или их частей, ширина которых при этом должна мало отличаться от ширины ближайших заданных интерференционных полос.

Известная теорема о сходимости рядов Фурье связывает скорость сходимости с числом m непрерывных производных представляемой рядом функции, $S_k = o(1/k^{m+1})$ при $k \rightarrow \infty$. Из этого выражения становится ясно, что при продолжении функции нежелательны скачки между известными фрагментами $U(x)$ и полученными в результате продолжения фрагментами $\tilde{U}(x)$. Но эта ситуация характерна для итерационного метода продолжения [29, 35]. Либо проводится значительное число итераций, и при наличии сходимости получается плавное сопряжение фрагментов, либо из-за недостатка времени осуществляется всего несколько итераций, и получается разрыв между функциями $\tilde{U}(x)$ и $U(x)$ в точках их сопряжения.

В статье [37] описывается техника дополнения интерференционных полос в отдельных сечениях интерферограммы синусоидой, начальная фаза, частота и амплитуда которой определяются исходя из известных фрагментов. По этому методу необходимо определять положения экстремальных точек в сечении.

Но в условиях действия шума эта операция некорректна и гладкость продолжения не реализуется.

Для дополнения сечения выберем в качестве модели его собственные фрагменты. При этом минимизируем на множестве отсчетов сечения некоторые функционалы гладкости.

Рассмотрим один из возможных алгоритмов реализации этого метода. Продолжение будем строить путем сдвига фрагментов сечения за пределы области определения. Пусть продолжаемое сечение определено на интервале $[1, n]$. Вначале продолжим правый край сечения на некоторое число отсчетов r , которое обеспечивает минимум функционала

$$L_r = \sum_{i=n-k+1}^n |U(i) - U(i-r)| = \min. \quad (28)$$

Затем продолжим левый край сечения на l отсчетов по условию

$$L_l = \sum_{i=1}^n |U(i) - U(i+l)| = \min. \quad (29)$$

Поиск минимума может осуществляться простым перебором. Одновременно минимизируются расстояния как между двумя функциями, так и между их разностями $-\Delta U(i)$, например при $k = 2$ получим:

$$\begin{aligned} \min &= |U(l) - U(l+1)| + |U(2) - U(l+2)| > \\ &> |U(2) - U(1)| - |U(l+2) - U(l+1)| = \\ &= |\Delta U(1) - \Delta U(l+1)|. \end{aligned} \quad (30)$$

Вновь образовавшиеся края для получения периодической функции сопрягаются между собой аналогично, но функционал гладкости будет зависеть от двух параметров

$$L_{pq} = \sum_{i=1}^n |U(r+i+p-1) - U(i+q)|. \quad (31)$$

Также перебором находится условный минимум L_{pq} , такой, что полная длина продолженного сечения $n_c = n + l + p + q$ принадлежит множеству чисел P_{10} .

Поэтому достаточная плотность этих чисел в интервале $[n, 3n]$, $r \in n_c$, является принципиальной, от нее зависит эффективность метода. Ниже приведены числа для наиболее употребительного диапазона от 2^7 до 2^8 .

128 130 132 133 135 136 138 140 143 144 147
 150 152 153 154 156 160 161 162 165 168 169
 170 171 175 176 180 182 184 187 189 190 192
 195 196 198 200 204 207 208 209 210 216 220
 224 225 228 234 240 242 243 245 250 252 256

Числа из множества P_{10} экспоненциально возрастают при линейном увеличении их номера. Следовательно, также уменьшается возможность продолжения на определенный интервал, начиная с некоторого номера. Возможность продолжения можно определить отношением количества чисел из P_{10} в заданном интервале к количеству чисел натурального ряда в

этом интервале (рис. 4). На рис. 5 показан пример продолжения сечения сигнала по описанному здесь алгоритму.

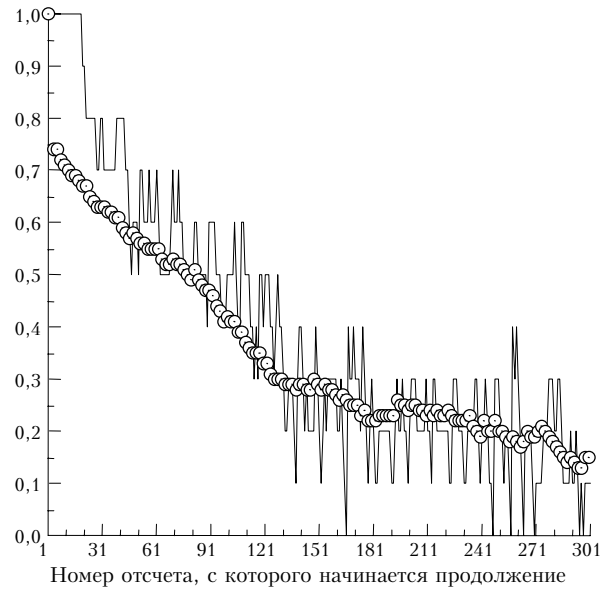


Рис. 4. Возможность продолжения сечения интерферограммы на десять отсчетов (ломаная линия), на сто отсчетов (значки) при разложении значения размерности массива на простые множители $\{1, 2, 3, 5, 7, 11, 13, 17, 19, 23\}$ для алгоритма быстрого преобразования Фурье

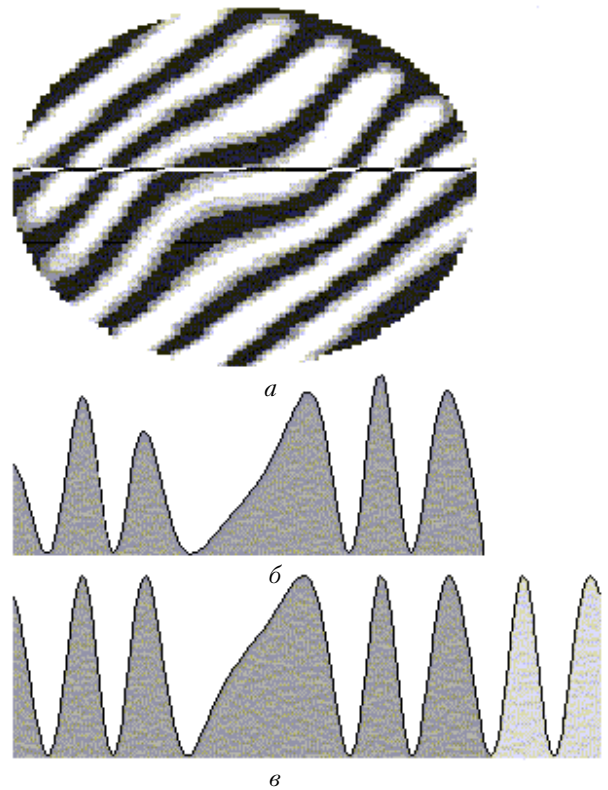


Рис. 5. Оптимальное продолжение сечения интерферограммы: *а* – интерферограмма, сечение выделено обратным контрастом; *б* – сечение интерферограммы; *в* – продолженное сечение, светлым тоном показано продолжение, полиномиальная модуляция устранена

Из приведенного описания метода продолжения можно сделать вывод, что он соответствует введенному выше критерию оптимальности при фиксации k , n , при $n_c \in P_{10}$ и в классе функций продолжаемого сечения. Численные эксперименты показали, что предложенный метод продолжения сечений интерферограммы за пределы области определения решает поставленную задачу и существенно уменьшает краевые эффекты у восстановленной фазы.

Выводы

Было проанализировано состояние проблемы определения амплитуды и фазы действительного сигнала и выделены две важные задачи. Это условие существования двухполосного спектра у действительного сигнала, которое является также условием, при котором возможно введение аналитического сигнала, и оптимальное вычисление преобразования Гильберта, что необходимо для численной реализации. В результате получены следующие результаты.

– Сформулировано условие дисперсионной причинности, использующее понятие эффективной ширины. Установлено, что монотонность фазы обеспечивает выполнение условия дисперсионной причинности для преобразования Фурье комплексного сигнала с постоянной амплитудой.

– Введен комплексный эффективный сигнал, который может быть применен при отсутствии строгой двухполосности у спектра. Разработан итерационный метод для его реализации.

– Для сигналов с монотонной фазой и постоянной амплитудой предложен метод демодуляции, не требующий построения комплексного сигнала, метод применим и для бинарных сигналов.

– Разработан алгоритм вычисления преобразования Гильберта, основанный на оптимальном периодическом продолжении преобразуемой функции по критерию минимума ширины спектра Фурье путем сдвига фрагментов функции за пределы области ее определения.

1. А.с. СССР № 1024746. Интерференционный датчик волнового фронта. Витриченко Э.А., Пушиной Л.А., Тартаковский В.А. Заявлено 18.02.82. Опубликовано 23.06.83. Бюллетень № 23.
2. Вакман Д.Е., Вайнштейн Л.А. Амплитуда, фаза, частота – основные понятия теории колебаний // Успехи физ. наук. 1977. Т. 123. Вып. 4. С. 657–682.
3. Вайнштейн Л.А., Вакман Д.Е. Разделение частот в теории колебаний и волн. М.: Наука, 1983. 287 с.
4. Vakman D. On the analytic signal, the Teager–Kaiser energy algorithm, and other methods for defining amplitude and frequency // IEEE Trans. Signal Process. 1996. V. 44. № 4. P. 791–797.
5. Vakman D. Analytic Waves // Int. J. Theor. Phys. 1997. V. 36. № 1. P. 227–247.
6. Cohen L., Loughlin P., Vakman D. On an ambiguity in the definition of the amplitude and phase of a signal // Signal Proc. 1999. V. 79. P. 301–307.
7. Витриченко Э.А., Пушиной Л.А., Тартаковский В.А. Интерференционный контроль оптики на основе диспер-

- сионных соотношений для логарифма аналитического сигнала // Докл. АН СССР. 1983. Т. 268. № 1. С. 91–95.
8. Витриченко Э.А., Лукин В.П., Пушиной Л.А., Тартаковский В.А. Проблемы оптического контроля. Новосибирск: Наука, 1990. 351 с.
9. Золотарев И.Д. Проблема «амплитуда, фаза, частота» и ее решение в радиотехнике // Техн. радиосвязи. 1997. № 3. С. 3–10.
10. Коржик В.И. Огибающая сигнала и некоторые ее свойства // Радиотехника. 1968. Т. 23. № 4. С. 1–6.
11. Левин Б.Я. Распределение корней целых функций. М.: Гостехиздат, 1956. 583 с.
12. Нуссенцвейг Х.М. Причинность и дисперсионные соотношения. М.: Мир, 1976. 461 с.
13. Папулис А. Теория систем и преобразований в оптике. М.: Мир. 1971. 495 с.
14. Сороко Л.М., Стриж Т.А. Спектральные преобразования на ЭВМ. Дубна: ОИЯИ, 1972. 135 с.
15. Тартаковский В.А. Программа дискретного преобразования Гильберта // ГосФАП. Алгоритмы и программы. 1978. Т. 6 (32). № 50. П003869.
16. Тартаковский В.А. О продолжении интерферограмм за пределы области определения // Оптика атмосф. и океана. 1993. Т. 6. № 12. С. 1581–1587.
17. Тартаковский В.А., Майер Н.Н. Дислокации фазы и минимально-фазовое разложение волновой функции // Оптика атмосф. и океана. 1995. Т. 8. № 3. С. 448–454.
18. Тартаковский В.А. Fringe pattern evaluation and Analytic Signal theory // Akademie Verlag Berlin. Optical Metrology Series. 1997. P. 84–91.
19. Тартаковский В.А. Определение фазы оптической волны и многомерный аналитический сигнал // Оптика атмосф. и океана. 1997. Т. 10. № 3. С. 301–315.
20. Тартаковский В.А., Майер Н.Н. Dispersion relation for real-plane zeros as a concept of wavefront measurement // Appl. Opt. 1998. V. 37. № 33. P. 7689–7697.
21. Финк Л.М. Соотношения между спектром и мгновенной частотой сигнала // Пробл. передачи информации. 1966. Т. II. Вып. 4. С. 65–73.
22. Хургии Я.И., Яковлев В.П. Прогресс в Советском Союзе в области теории финитных функций и ее применений в физике и технике // ТИИЭР. 1977. Т. 65. № 7. С. 16–45.
23. Хургии Я.И., Яковлев В.П. Финитные функции в физике и технике. М.: Наука, 1971. 408 с.
24. Bedrosian E.A. product theorem for Hilbert transforms // Proc. IEEE. 1963. V. 51. № 5. P. 868–869.
25. Čizek V. Numerische Hilbert-Transformation // Práce ústavu radiotechniky a elektroniky. Československé akademie věd. 1961. № 11. P. 56.
26. Cooley J.W. and Tukey J.W. An algorithm for the machine calculation of complex Fourier series // Mathematical Computation. 1965. V. 19. P. 297–301.
27. Gabor D. Theory of communication // J. of IEE. 1946. V. 93. Pt. 3. P. 429–441.
28. Harris F.J. Использование окон при гармоническом анализе методом дискретного преобразования Фурье // ТИИЭР. 1978. Т. 66. № 1. С. 60–96.
29. Huang T.C., Sanz J.L.C., Fan H. et al. Numerical comparison of several algorithms for band-limited signal extrapolation // Appl. Opt. 1984. V. 23. № 2. P. 307–317.
30. Kak S.C. The discrete Hilbert transform // Proc. IEEE. 1969. № 4. P. 585–586.
31. Lai G., Yataqai T. Use of the fast Fourier transform method for analyzing linear and equispaced Fizeau fringes // Appl. Opt. 1994. V. 33. № 25. P. 5935–5940.

32. *Mandel L.* Complex representation of optical fields in coherence theory // *J. Opt. Soc. Amer.* 1967. V. 57. № 5. P. 613–617.
33. *Requicha A.* Нули целых функций: Теория и инженерные приложения // *ТИИЭР.* 1980. Т. 68. № 3. С. 5–30.
34. *Rice S.O.* Огибающие узкополосных сигналов // *ТИИЭР.* 1982. Т. 70. № 7. С. 5–13.
35. *Shafer R.U., Mersero Z.M., Richards M.A.* Итерационные алгоритмы восстановления сигналов при наличии ограничений // *ТИИЭР.* 1981. Т. 69. № 4. С. 33–55.
36. *Singleton R.C.* An Algorithm for computing the mixed radix fast Fourier transform // *IEEE Trans. audio and electroacoustics.* 1969. V. AU-17. № 2. P. 93–103.
37. *Spik A.* Two-dimensional phase decoding from bounded fringe patterns by using the Fourier-transform method // *Opt. Appl.* 1987. V. XVII. № 4. P. 349–354.
38. *Takeda M., Ina H., Kobayashi S.* Fourier-transform method of fringe-pattern analysis for computer-based topography and interferometry // *Appl. Opt.* 1982. V. 21. № 1. P. 157–160.
39. *Wolf E., Agarwal G.S.* Relation between the statistical representations of real and associated complex fields in optical coherence theory // *J. Math. Phys.* 1972. V. 13. № 11. P. 1759–1764.

Valery A. Tartakovsky. **Causality and demodulation of optical signals with the monotonous phase.**

The-state-of-the-art of definition of an envelope and a phase of signals as applied to analysis of fringe patterns like interferogram is discussed. Two problems having both theoretical and practical significance were considered, namely, the existence conditions of two-band spectrum at real signals and optimal calculation of the Hilbert transform. Constructive results are presented.