

Преобразование  $X(z)$  дискретизировано в  $N$  равностоящих точках на единичной окружности:

$$\hat{X}_p(k) = \hat{X}\left(e^{i\frac{2\pi}{N}k}\right), \quad 0 \leq k \leq N-1.$$

Используя ДПФ, вычисляем

$$\hat{x}_p(n) = \frac{1}{N} \sum_{k=0}^{N-1} \hat{X}_p(k) e^{i\frac{2\pi}{N}kn}, \quad 0 \leq n \leq N-1,$$

что может служить аппроксимацией комплексного кепстра.

а) Выразить  $X_p(k)$  через действительный кепстр  $\hat{x}(m)$ .

б) Подставить выражение п.а) в обратное преобразование Фурье для

$$\hat{x}_p(n) \text{ и показать, что } \hat{x}_p(n) = \sum_{r=-\infty}^{\infty} \hat{x}(n+rN).$$

7.7. Рассмотрим последовательность

$$x(n) = \delta(n) + \alpha\delta(n - N_p).$$

а) Определить комплексный кепстр. Изобразить ваш результат.

б) Изобразить кепстр  $c(n)$  для  $x(n)$ .

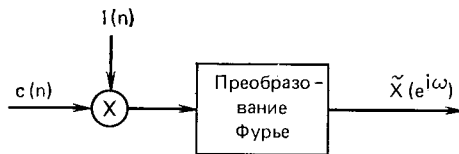


Рис. 3.7.1

в) Предположим, что по (7.30) вычислено приближение  $x_p(n)$ . Изобразить  $x_p(n)$  для  $0 \leq n \leq N-1$  в случае  $N_p = N/6$ . Что, если  $N$  не делится нацело на  $N_p$ ?

г) Повторить п.в) для кепстральной аппроксимации  $c_p(n)$  при  $0 \leq n \leq N-1$  с использованием (7.33).

д) Если наибольшее значение кепстральной аппроксимации  $c_p(n)$  используется для определения  $N_p$ , то как велико должно быть  $N$  для того, чтобы избежать ошибок?

7.8. Для сглаживания логарифма модуля спектра сигнала его кепстр часто взвешивают и преобразование Фурье имеет вид рис. 3.7.1.

а) Написать выражение, связывающее  $X(e^{j\omega})$  с  $\log|X(e^{j\omega})|$  и  $L(e^{j\omega})$ , где  $L(e^{j\omega})$  — преобразование Фурье  $l(n)$ .

б) Какое кепстральное окно следует использовать для сглаживания функции  $\log|X(e^{j\omega})|$ ?

в) Сравнить применение прямоугольного окна и окна Хемминга в качестве кепстральных окон.

г) Какова должна быть протяженность кепстрального окна и почему?

7.9. Рассмотрим сегмент вокализованного сигнала, который можно представить в виде  $s(m) = p(m) * h_v(m)$ , где  $p(m) = \sum_{r=-\infty}^{\infty} \delta(m-rN_p)$ . При вычислении

комплексного кепстра (или кепстра) первый шаг заключается в умножении  $s(m)$  на окно  $w(m)$  для выделения сегмента  $x_n(m) = s(m)w(n-m)$  входных данных для гомоморфной обработки.

а) Определить условия, при которых можно аппроксимировать  $x_n(m)$  в виде  $x(m) = p_n(m) * h_v(m)$ , где  $p_n(m) = p(m)w(n-m)$ .

б) Для специального случая  $n=0$  определить преобразование  $p_0(n)$  через  $z$ -преобразование  $w(m)$ .

в) Выразить комплексный кепстр  $\hat{p}_0(m)$  через  $\hat{w}(m)$ .

7.10. В задаче 7.9 показано, что периодичность взвешенного сегмента вокализованной речи может быть приближенно представлена выражением  $p_n(m) = p(m)w(n-m)$ , где  $p(m) = \sum_{r=-\infty}^{\infty} \delta(m-rN_p)$ . В этой задаче исследуется влияние

положения окна на комплексный кепстр  $\hat{p}_n(m)$ . Предположим, что имеется окно Хемминга вида

$$w(m) = \begin{cases} 0,54 - 0,46 \cos(2\pi m/(2N_p)), & 0 \leq m \leq 2N_p; \\ 0, & \text{в противном случае.} \end{cases}$$

а) Изобразить  $p_n(m)$  как функцию  $m$  для  $n=3N_p/4; 9N_p/8; 5N_p/4; 3N_p/2$ .

б) Для каждого из перечисленных выше случаев составить выражение для  $p_n(m)$  и показать, что соответствующее  $z$ -преобразование имеет вид  $P_n(z) = \alpha_1 z^{N_p} + \alpha_2 + \alpha_3 z^{-N_p}$ .

в) Для каждого из перечисленных выше случаев определить и изобразить комплексный кепстр (указание: использовать разложение в ряд для  $\log[P_n(z)]$ ). Опустить члены вида  $\log[z^{\pm N_p}]$ .

г) Для какого положения окна справедливы следующие утверждения:

последовательность  $p_n(m)$  минимально-фазовая;

последовательность  $p_n(m)$  максимально-фазовая;

первый кепстральный пик максимален;

первый кепстральный пик минимален.

д) Как изменятся ваши ответы на перечисленные выше вопросы, если окно удлинится? Укоротится?

7.11. Преобразование сигнала  $x(n)$  определяется как

$$X(z) = \sum_{n=0}^{N-1} x(n) z^{-n}.$$

Вычислим  $X(z)$  в последовательности точек  $z_k = AW^{-k}, k=0, 1, \dots, M-1$ , где  $A$  и  $W$  — произвольные комплексные целые числа. Если сделать простую подстановку  $nk = [n^2 + k^2 - (k-n)^2]/2$ , то  $X(z_k)$  можно записать в виде

$$X(z_k) = P(k) \sum_{n=0}^{N-1} y(n) g(k-n),$$

т. е.  $X(z_k)$  — свертка  $y(n)$  и  $g(n)$ .

а) Определить  $P(k)$ ,  $y(n)$  и  $g(n)$  через  $x(n)$ ,  $A$  и  $W$ .

б) Изобразить точки  $z_k$  на  $z$ -плоскости.

в) Можете ли Вы предложить способ применения БПФ для вычисления приведенного выше выражения?

## 8

# Кодирование речевых сигналов на основе линейного предсказания

## 8.0. Введение

Линейное предсказание является одним из наиболее эффективных методов анализа речевого сигнала. Этот метод становится доминирующим при оценке основных параметров речевого сигнала, таких, как, например, период основного тона, форманты, спектр, функция площади речевого тракта, а также при сокращенном представлении речи с целью ее низкоскоростной передачи и экономного хранения. Важность метода обусловлена высокой точностью получаемых оценок и относительной простотой вычислений. В данной главе излагаются основные положения метода линейного предсказания и приводятся рекомендации по его практическому использованию.

Основной принцип метода линейного предсказания состоит в том, что текущий отсчет речевого сигнала можно аппроксимировать линейной комбинацией предшествующих отсчетов. Коэффициенты предсказания при этом определяются однозначно минимизацией среднего квадрата разности между отсчетами речевого сигнала и их предсказанными значениями (на конечном интервале). Коэффициенты предсказания — это весовые коэффициенты, используемые в линейной комбинации.

Основные положения метода линейного предсказания хорошо согласуются с моделью речеобразования, рассмотренной в гл. 3, где показано, что речевой сигнал можно представить в виде сигнала на выходе линейной системы с переменными во времени параметрами, возбуждаемой квазипериодическими импульсами (в пределах вокализованного сегмента) или случайным шумом (на невокализованном сегменте). Метод линейного предсказания позволяет точно и надежно оценить параметры этой линейной системы с переменными коэффициентами.

Линейное предсказание уже обсуждалось в гл. 5 в связи с решением задачи квантования речи. Там предполагалось, что метод линейного предсказания можно применять для сокращения объема цифрового речевого сигнала.

Математические основы метода, используемого в адаптивном предсказателе высокого порядка при АРИКМ-кодировании, по существу совпадают с рассматриваемыми в этой главе. При АРИКМ-кодировании основная задача состоит в построении предсказателя, который минимизировал бы ошибку предсказания и, следовательно, шум квантования. В данной главе с более общих позиций будет показано, как основные идеи линейного предсказания приводят к ряду методов, которые можно использовать при оценке параметров речевых моделей.

Идеи и методы линейного предсказания довольно давно обсуждаются в технической литературе. Эти идеи используются в теориях автоматического управления и информации, где их называют методами оценивания систем, или методами идентификации систем. Под термином «идентификация» понимаются методы линейного предсказания (ЛП), основанные на оценивании параметров, однозначно описывающих систему при условии, что ее передаточная функция является полусной. Применительно к обработке речевых сигналов методы линейного предсказания означают ряд сходных формулировок задачи моделирования речевого сигнала [1—18]. Эти формулировки часто отличаются в исходных предположениях. Иногда они сводятся к различным методам вычисления, используемым для оценки коэффициентов предсказания. Так, применительно к речевым сигналам существуют следующие методы вычисления (часто равноценные): ковариационный [3], автокорреляционный [1, 2, 9], лестничного фильтра [11, 12], обратной фильтрации [1], оценки спектра [12], максимального правдоподобия [4, 6] и скалярного произведения [1]. В этой главе подробно рассматриваются сходства и различия трех первых из перечисленных выше методов, поскольку остальные подходы равноценны одному из этих трех.

Целесообразность использования линейного предсказания обусловлена высокой точностью описания речевого сигнала с помощью модели. Поэтому большая часть данной главы содержит методы оценивания различных параметров речевого сигнала с помощью линейного предсказания. Далее обсуждается ряд типичных примеров применения методов линейного предсказания, а в гл. 9 представлены ряд задач, для успешного решения которых также применяются методы линейного предсказания.

## 8.1. Методы анализа на основе линейного предсказания

В книге неоднократно использовалась основная модель речеобразования в дискретном времени, предложенная в гл. 3. На рис. 8.1 эта модель представлена в форме, наиболее удобной для решения задач линейного предсказания. В этом случае общий спектр, обусловленный излучением, речевым трактом и возбуж-

дением, описывается с помощью линейной системы с переменными параметрами и передаточной функцией

$$H(z) = \frac{S(z)}{U(z)} = \frac{G}{1 - \sum_{k=1}^p a_k z^{-k}} \quad (8.1)$$

Эта система возбуждается импульсной последовательностью для вокализованных звуков речи и шумом для невокализованных. Таким образом, модель имеет следующие параметры: классификатор вокализованных и невокализованных звуков, период основного тона для вокализованных сегментов, коэффициент усиления  $G$  и коэффициенты  $\{a_k\}$  цифрового фильтра. Все эти параметры, разумеется, медленно изменяются во времени.

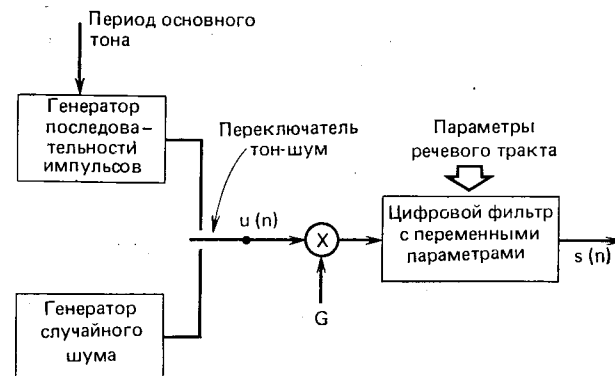


Рис. 8.1. Структурная схема упрощенной модели речеобразования

Определение периода основного тона и классификация тон/шум могут быть осуществлены на основе использования ряда методов, уже обсуждавшихся в данной книге, или с помощью рассматриваемых ниже методов линейного предсказания. Как отмечалось в гл. 3, для вокализованных звуков хорошо подходит модель, содержащая только полюсы в своей передаточной функции (чисто полюсная), но для носовых и фрикативных звуков требуется учитывать и нули. Однако из дальнейшего будет ясно, что если порядок  $p$  модели достаточно велик, то полюсная модель позволяет получить достаточно точное описание почти для всех звуков речи. Главное достоинство этой модели заключается в том, что как параметр  $G$ , так и коэффициенты можно оценить непосредственно с использованием очень эффективных с вычислительной точки зрения алгоритмов.

Для системы рис. 8.1 отсчет речевого сигнала  $s(n)$  связан с сигналом возбуждения  $u(n)$  простым разностным уравнением

$$s(n) = \sum_{k=1}^p a_k s(n-k) + Gu(n) \quad (8.2)$$

Линейный предсказатель с коэффициентами  $\alpha_k$  определяется как система, на выходе которой имеем

$$s(n) = \sum_{k=1}^p \alpha_k s(n-k). \quad (8.3)$$

Такие системы использовались в гл. 5 для уменьшения дисперсии погрешности предсказания. Системная функция предсказателя  $p$ -го порядка представляет собой полином вида

$$P(z) = \sum_{k=1}^p \alpha_k z^{-k}. \quad (8.4)$$

Погрешность предсказания определяется как

$$e(n) = s(n) - \tilde{s}(n) = s(n) - \sum_{k=1}^p \alpha_k s(n-k). \quad (8.5)$$

Из (8.5) видно, что погрешность предсказания представляет собой сигнал на выходе системы с передаточной функцией

$$A(z) = 1 - \sum_{k=1}^p \alpha_k z^{-k}. \quad (8.6)$$

Сравнение (8.2) и (8.5) показывает, что если сигнал точно удовлетворяет модели (8.2) и  $\alpha_k = a_k$ , то  $e(n) = Gu(n)$ . Таким образом, *фильтр погрешности предсказания*  $A(z)$  является *обратным фильтром* для системы  $H(z)$ , соответствующей уравнению (8.1), т. е.

$$H(z) = G/A(z). \quad (8.7)$$

Основная задача анализа на основе линейного предсказания заключается в непосредственном определении параметров  $\{\alpha_k\}$  по речевому сигналу с целью получения хороших оценок его спектральных свойств путем использования уравнения (8.7). Вследствие изменения свойств речевого сигнала во времени коэффициенты предсказания должны оцениваться на коротких сегментах речи. Основным подходом является определение параметров предсказания таким образом, чтобы минимизировать дисперсию погрешности на коротком сегменте сигнала. При этом *предполагается*, что полученные параметры являются параметрами системной функции  $H(z)$  в модели речеобразования.

То, что подобный подход приводит к полезным результатам, возможно, не сразу очевидно, но его полезность будет неоднократно подтверждена различными способами. Во-первых, пусть  $\alpha_k = a_k$ , тогда  $e(n) = Gu(n)$ . Для вокализованной речи это означает, что  $e(n)$  будет состоять из последовательности импульсов, т. е.  $e(n)$  будет весьма мало почти все время. Поэтому в данном случае минимизация погрешности предсказания позволит получить требуемые коэффициенты. Другой повод, приводящий к тому же

подходу, вытекает из того, что даже если сигнал формируется системой (8.2) с постоянными во времени параметрами, которая возбуждается либо единичным импульсом, либо белым шумом, то можно показать, что коэффициенты предсказания, найденные по критерию минимизации среднего квадратического значения погрешности (в каждый момент времени), совпадают с коэффициентами в (8.2). Третьей, весьма важной для практики причиной является то, что подобная минимизация приводит к линейной системе уравнений, решение которых сравнительно легко приводит к получению параметров предсказания. Кроме того, полученные параметры, как это будет ясно из дальнейшего, составляют весьма плодотворную основу для точного описания сигнала.

Кратковременная энергия погрешности предсказания:

$$E_n = \sum_m e_n^2(m) = \quad (8.8)$$

$$= \sum_m (s_n(m) - \tilde{s}_n(m))^2 = \quad (8.9)$$

$$= \sum_m \left[ s_n(m) - \sum_{k=1}^p \alpha_k s_n(m-k) \right]^2, \quad (8.10)$$

где  $s_n(m)$  — сегмент речевого сигнала, выбранный в окрестности отсчета  $n$ , т. е.

$$s_n(m) = s(m+n). \quad (8.11)$$

Пределы суммирования справа в (8.8) — (8.10) пока не определены, но поскольку предполагается использовать концепции кратковременного анализа, то эти пределы всегда предполагаются конечными. Кроме того, для получения среднего значения необходимо разделить полученный результат на длину речевого сегмента. Однако эти константы несущественны с точки зрения решения системы линейных уравнений и поэтому далее опускаются. Параметры  $\alpha_k$  можно получить, минимизируя  $E_n$  в (8.10) путем вычисления  $\frac{\partial E_n}{\partial \alpha_i} = 0$ ,  $i = 1, 2, \dots, p$ , что приводит к системе уравнений

$$\sum_m s_n(m-i) s_n(m) = \sum_{k=1}^p \hat{\alpha}_k \sum_m s_n(m-i) s_n(m-k), \quad 1 \leq i \leq p, \quad (8.12)$$

где  $\hat{\alpha}_k$  — значения  $\alpha_k$ , минимизирующие  $E_n$  (поскольку значение  $\alpha_k$  — единственное, далее знак  $\wedge$  опускается и за величину, минимизирующую  $E_n$ , принимается  $\alpha_k$ ). Если ввести определение

$$\Phi_n(i, k) = \sum_m s_n(m-i) s_n(m-k), \quad (8.13)$$

тогда (8.12) можно переписать в более компактном виде:

$$\sum_{k=1}^p \alpha_k \varphi_n(i, k) = \varphi_n(i, 0), \quad i = 1, 2, \dots, p. \quad (8.14)$$

Эта система из  $p$  уравнений с  $p$  неизвестными может быть решена достаточно эффективным способом для получения неизвестных коэффициентов предсказания, минимизирующих средний квадрат погрешности предсказания на сегменте  $s_n(m)$ <sup>1</sup>. Используя (8.10) и (8.12), можно показать, что средняя квадратическая погрешность предсказания имеет вид

$$E_n = \sum_m s_n^2(m) - \sum_{k=1}^p \alpha_k \sum_m s_n(m) s_n(m-k) \quad (8.15)$$

и, используя (8.14), можно выразить  $E_n$  в виде

$$E_n = \varphi_n(0, 0) - \sum_{k=1}^p \alpha_k \varphi_n(0, k). \quad (8.16)$$

Таким образом, общая погрешность предсказания состоит из двух слагаемых, одно из которых является постоянным, а другое — зависит от коэффициентов предсказания.

Для решения системы уравнений относительно коэффициентов предсказания следует первоначально вычислить величины  $\varphi_n(i, k)$ ,  $1 \leq i \leq p$  и  $0 \leq k \leq p$ . Только после этого можно переходить к решению (8.14) и получению оценок  $\alpha_k$ . Таким образом, принципиально анализ на основе линейного предсказания очень простой. Однако подробности, связанные с вычислением  $\varphi_n(i, k)$  и последующим решением системы уравнений, являются достаточно запутанными и нуждаются в дальнейшем обсуждении.

Хотя пределы суммирования в (8.8) — (8.10) и (8.12) не определены, заметим, что в (8.12) они совпадают с соответствующими пределами в (8.8) — (8.10). Как было установлено, для кратковременного анализа соответствующие пределы должны охватывать конечный интервал. Имеется два подхода к этому вопросу, и, как это будет ясно из дальнейшего, в зависимости от пределов суммирования и выбора сегмента  $s_n(m)$  различают два метода линейного предсказания.

### 8.1.1. Автокорреляционный метод [1, 2, 5]

Один из способов определения пределов в (8.8) — (8.10) и (8.12) основан на предположении, что сигнал равен нулю вне интервала  $0 \leq m \leq N-1$ . Это удобно записать в виде

$$s_n(m) = s(m+n)w(m), \quad (8.17)$$

<sup>1</sup> Очевидно, что  $\alpha_k$  зависит от  $n$  (момент времени, для которого получена оценка), хотя эта зависимость не показана в явном виде. Далее индекс времени при  $E_n$ ,  $s_n(m)$  и  $\varphi_n(i, k)$  опускается, если это не вызовет затруднений.

где  $w(m)$  — окно конечной длительности (например, окно Хемминга), равное нулю вне интервала.

Значение этого предположения при решении вопроса о пределах суммирования в выражении для  $E_n$  можно рассмотреть на примере соотношения (8.5). Очевидно, что если  $s_n(m)$  отличен от нуля только на интервале  $0 \leq m \leq N-1$ , то соответствующая погрешность предсказания  $e_n(m)$  для предсказателя порядка  $p$  будет отлична от нуля на интервале  $0 \leq m \leq N-1+p$ . В этом случае  $E_n$  имеет вид

$$E_n = \sum_{m=0}^{N+p-1} e_n^2(m). \quad (8.18)$$

С другой стороны, легко показать, что пределы суммирования можно распространить на все ненулевые значения на интервале от  $-\infty$  до  $+\infty$  [2].

Возвращаясь к (8.5), можно отметить, что погрешность предсказания будет, вероятно, большой в начале интервала (т. е.  $0 \leq m \leq p-1$ ), поскольку мы пытаемся предсказать сигнал по отсчетам, которые приравняли нулю. Очевидно, что погрешность будет большой и в конце интервала (т. е.  $N \leq m \leq N+p-1$ ), поскольку здесь мы предсказываем нулевые значения по ненулевым. Поэтому в качестве окна  $w(m)$  в уравнении (8.17) используется окно, которое стремится к нулю на концах интервала.

Пределы при вычислении  $\varphi_n(i, k)$  в (8.13) совпадают с пределами (8.18). Но, поскольку  $s_n(m)$  равно нулю вне интервала  $0 \leq m \leq N-1$ , легко показать, что

$$\varphi_n(i, k) = \sum_{m=0}^{N+p-1} s_n(m-i) s_n(m-k), \quad 1 \leq i \leq p, \quad 0 \leq k \leq p, \quad (8.19a)$$

можно выразить в виде

$$\varphi_n(i, k) = \sum_{m=0}^{N-1-(i-k)} s_n(m) s_n(m+i-k), \quad 1 \leq i \leq p, \quad 0 \leq k \leq p. \quad (8.19b)$$

Легко видеть, что в данном случае  $\varphi_n(i, k)$  совпадает с кратковременной автокорреляционной функцией сигнала (4.30), вычисленной для  $(i-k)$ . Это означает, что

$$\varphi_n(i, k) = R_n(i-k), \quad (8.20)$$

где

$$R_n(k) = \sum_{m=0}^{N-1-k} s_n(m) s_n(m+k). \quad (8.21)$$

Вычисление  $R_n(k)$  детально обсуждалось в § 4.6 и здесь не рассматривается. Поскольку  $R_n(k)$  — четная функция, то

$$\varphi_n(i, k) = R_n(|i-k|), \quad i = 1, 2, \dots, p, \quad k = 0, 1, \dots, p. \quad (8.22)$$

Таким образом, (8.14) можно представить в виде

$$\sum_{k=1}^p \alpha_k R_n(|i-k|) = R_n(i) \quad 1 \leq i \leq p. \quad (8.23)$$

Аналогично минимальный средний квадрат погрешности предсказания

$$E_n = R_n(0) - \sum_{k=1}^p \alpha_k R_n(k). \quad (8.24)$$

Систему уравнений (8.23) можно записать в матричной форме:

$$\begin{pmatrix} R_n(0) & R_n(1) & R_n(2) & \dots & R_n(p-1) \\ R_n(1) & R_n(0) & R_n(1) & \dots & R_n(p-2) \\ R_n(2) & R_n(1) & R_n(0) & \dots & R_n(p-3) \\ \dots & \dots & \dots & \dots & \dots \\ R_n(p-1) & R_n(p-2) & R_n(p-3) & \dots & R_n(0) \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \dots \\ \alpha_p \end{pmatrix} = \begin{pmatrix} R_n(1) \\ R_n(2) \\ R_n(3) \\ \dots \\ R_n(p) \end{pmatrix}. \quad (8.25)$$

Матрица размером  $p \times p$  является теплицевой, т. е. симметричной и такой, что элементы на любой диагонали равны между собой. Это особое свойство будет использовано в § 8.3 для получения эффективного алгоритма решения уравнений (8.23).

### 8.1.2. Ковариационный метод [3]

Другой основной подход к определению сегмента речевого сигнала и пределов суммирования заключается в том, что фиксируется интервал, на котором вычисляется средний квадрат погрешности, и рассматривается влияние этого обстоятельства на вычисление  $\varphi_n(i, k)$ . Другими словами, если определить

$$E_n = \sum_{m=0}^{N-1} e_n^2(m), \quad (8.26)$$

то  $\varphi_n(i, k)$  выражается формулой

$$\varphi_n(i, k) = \sum_{m=0}^{N-1} s_n(m-i) s_n(m-k), \quad 1 \leq i \leq p, 0 \leq k \leq p. \quad (8.27)$$

Изменив индекс суммирования, (8.27) можно выразить в виде

$$\varphi_n(i, k) = \sum_{m=-i}^{N-i-1} s_n(m) s_n(m+i-k), \quad 1 \leq i \leq p, 0 \leq k \leq p, \quad (8.28a)$$

или

$$\varphi_n(i, k) = \sum_{m=-k}^{N-k-1} s_n(m) s_n(m+k-i), \quad 1 \leq i \leq p, 0 \leq k \leq p. \quad (8.28b)$$

Полученные уравнения кажутся очень похожими на (8.19б), однако они имеют иные пределы суммирования. В (8.28) используется значение сигнала  $s_n(m)$  вне интервала  $0 \leq m \leq N-1$ . Действительно, для вычисления  $\varphi_n(i, k)$  для всех требуемых значений  $i$  и  $k$  необходимо использовать значения  $s_n(m)$  на интервале  $-p \leq m \leq N-1$ . Для того чтобы это не противоречило пределам суммирования в (8.26), в данном случае используются необходимые значения сигнала без ограничения последовательности отсчетов окном конечной длительности, уменьшающимся к концам интервала, как это имело место в автокорреляционном методе. Таким образом, здесь используются отсчеты и вне интервала  $0 \leq m \leq N-1$ . Очевидно, что данный метод похож на метод вычисления модифицированной автокорреляционной функции (см. гл. 4). Как указывалось в § 4.6, это приводит не к автокорреляционной, а к взаимокорреляционной функции между двумя очень сходными, но не одинаковыми сегментами речевого сигнала конечной длительности. Хотя различие между (8.28) и (8.19) сводится к небольшим вычислительным подробностям, система уравнений

$$\sum_{k=1}^p \alpha_k \varphi_n(i, k) = \varphi_n(i, 0), \quad i = 1, 2, \dots, p, \quad (8.29a)$$

обладает свойствами, которые значительно влияют на метод решения и свойства получаемого оптимального предсказателя. В матричной форме система уравнений имеет вид

$$\begin{pmatrix} \varphi_n(1,1) & \varphi_n(1,2) & \varphi_n(1,3) & \dots & \varphi_n(1,p) \\ \varphi_n(2,1) & \varphi_n(2,2) & \varphi_n(2,3) & \dots & \varphi_n(2,p) \\ \varphi_n(3,1) & \varphi_n(3,2) & \varphi_n(3,3) & \dots & \varphi_n(3,p) \\ \dots & \dots & \dots & \dots & \dots \\ \varphi_n(p,1) & \varphi_n(p,2) & \varphi_n(p,3) & \dots & \varphi_n(p,p) \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \dots \\ \alpha_p \end{pmatrix} = \begin{pmatrix} \varphi_n(1,0) \\ \varphi_n(2,0) \\ \varphi_n(3,0) \\ \dots \\ \varphi_n(p,0) \end{pmatrix} \quad (8.29b)$$

В этом случае, поскольку  $\varphi_n(i, k) = \varphi_n(k, i)$  [см. (8.28)], матрица размером  $p \times p$  является квазикорреляционной симметричной, но не теплицевой. Действительно, можно сказать, что диагональные элементы связаны соотношением

$$\varphi_n(i+1, k+1) = \varphi_n(i, k) + s_n(-i-1) s_n(-k-1) - s_n(N-1-i) s_n(N-1-k). \quad (8.30)$$

Метод анализа, основанный на изложенном выше способе вычисления  $\varphi_n(i, k)$ , известен как *ковариационный метод*, поскольку матрица обладает свойствами ковариационной матрицы [5]<sup>1</sup>.

<sup>1</sup> Эта терминология отличается от общепринятой, поскольку термин «ковариация» обычно означает корреляцию сигнала, из которого предварительно вычитается среднее значение.

### 8.1.3. Заключение

Показано, что в зависимости от определения сегмента анализируемого сигнала можно получить две различные системы уравнений. Для автокорреляционного метода сигнал взвешивается с использованием  $N$ -точечного окна и величины  $\varphi_n(i, k)$  получаются на основе кратковременной автокорреляционной функции. Полученная матрица корреляций является теплицевой и приводит к первой системе уравнений для параметров предсказания. При ковариационном методе сигнал предполагается известным на множестве значений  $-p \leq n \leq N-1$ . Никаких предположений о сигнале вне данного интервала не делается, поскольку только этот интервал необходим для вычислений. Полученная матрица корреляций в данном случае симметричная, но не теплицева. В результате два различных метода вычисления корреляции приводят к двум различным системам уравнений и к двум совокупностям коэффициентов предсказания с различными свойствами.

В последующих параграфах проводятся сравнение и противопоставление вычислительных процедур и результатов, даваемых обоими методами, наряду с рассматриваемыми ниже другими методами. Однако сначала рассмотрим определение коэффициента усиления  $G$  на рис. 8.1 на основе выражения для погрешности предсказания.

### 8.2. Вычисление коэффициента усиления модели [2]

Естественно ожидать, что коэффициент усиления  $G$  можно определить путем согласования энергии сигнала и линейно-предсказанных отсчетов. Это действительно верно, если сделать соответствующие предположения относительно сигнала возбуждения в модели с линейным предсказанием.

Постоянную  $G$  можно включить в сигнал возбуждения и ошибку предсказания (8.2) и (8.3)<sup>1</sup>. Сигнал возбуждения можно представить в виде

$$G u(n) = s(n) - \sum_{k=1}^p a_k s(n-k), \quad (8.31a)$$

при этом погрешность предсказания будет представлена в виде

$$e(n) = s(n) - \sum_{k=1}^p \alpha_k s(n-k). \quad (8.31b)$$

В случае, когда  $a_k = \alpha_k$ , т. е. коэффициенты предсказания совпадают с коэффициентами модели,

$$e(n) = G u(n), \quad (8.32)$$

<sup>1</sup> Отметим, что усиление также является функцией времени.

т. е. входной сигнал модели пропорционален погрешности предсказания с коэффициентом пропорциональности  $G$ ; подробнее свойства погрешности предсказания рассматриваются в § 8.5.

Поскольку (8.32) является приближенным (т. е. справедливо лишь при равенстве параметров модели и предсказателя), в общем случае определить  $G$  непосредственно по погрешности предсказания невозможно. Целесообразнее допустить, что энергия погрешности предсказания равна энергии сигнала возбуждения, т. е.

$$G^2 \sum_{m=0}^{N-1} u^2(m) = \sum_{m=0}^{N-1} e^2(m) = E_n. \quad (8.33)$$

В этом случае для определения  $G$  по каким-либо параметрам, например по  $a_k$  и коэффициентам корреляции, необходимы некоторые предположения относительно  $u(n)$ . Имеются два случая, для которых требуются соответствующие предположения. Для вокализованной речи естественно предположить, что  $u(n) = \delta(n)$ , т. е. возбуждение представляет собой единичный отсчет в нулевой момент времени<sup>1</sup>. Для справедливости этого предположения необходимо, чтобы линейный предсказатель с переменными во времени параметрами описывал как передаточную функцию речевого тракта, так и различие в форме между реальными импульсами возбуждения на вокализованном сегменте и единичными импульсами в модели. Для этого необходимо, чтобы порядок предсказателя был достаточным для описания как передаточной функции речевого тракта, так и эффекта возбуждения. Выбор порядка предсказателя будет рассмотрен в последующих разделах. Для невокализованных сегментов целесообразно предположить, что  $u(n)$  представляет собой стационарный белый шум с нулевым средним и единичной дисперсией.

Используя эти предположения, определим  $G$  на основе соотношения (8.33). Для вокализованных сегментов на входе имеется сигнал  $G\delta(n)$ . Если обозначить сигнал на выходе для этого случая через  $h(n)$  [поскольку в действительности это есть импульсная характеристика системы с передаточной функцией  $H(z)$ , как в (8.1)], получим соотношение

$$h(n) = \sum_{k=1}^p \alpha_k h(n-k) + G\delta(n). \quad (8.34)$$

Можно показать, что автокорреляционная функция для  $h(n)$  (см. задачу 8.1), определяемая как

$$\tilde{R}(m) = \sum_{n=0}^{\infty} h(n) h(m+n), \quad (8.35)$$

<sup>1</sup> Для справедливости этого предположения необходимо, чтобы интервал анализа совпадал с периодом основного тона.

удовлетворяет соотношениям

$$\tilde{R}(m) = \sum_{k=1}^p \alpha_k \tilde{R}(|m-k|), \quad m = 1, 2, \dots, p, \quad (8.36a)$$

$$\tilde{R}(0) = \sum_{k=0}^p \alpha_k \tilde{R}(k) + G^2. \quad (8.36b)$$

Поскольку (8.36) совпадает с (8.32), следовательно,

$$\tilde{R}(m) = R_n(m), \quad 1 \leq m \leq p. \quad (8.37)$$

Учитывая, что полные энергии сигнала  $[R(0)]$  и импульсной реакции  $[\tilde{R}(0)]$  должны быть равны, можно использовать (8.24), (8.33) и (8.34), чтобы получить

$$G^2 = R_n(0) - \sum_{k=1}^p \alpha_k (R_n(k) = E_n. \quad (8.38)$$

Интересно, что требование равенства и (8.37) приводят к тому, что первые  $p+1$  коэффициентов автокорреляции импульсной характеристики модели и сигнала должны совпадать.

В случае некогерентной речи корреляционная функция определяется статистическим усреднением. Предполагается, что сигнал возбуждения — белый шум с нулевым средним и единичной дисперсией, т. е.

$$E[u(n)u(n-m)] = \delta(m). \quad (8.39)$$

Если возбудить систему случайным процессом  $Gu(n)$  и обозначить процесс на выходе через  $g(n)$ , то

$$g(n) = \sum_{k=1}^p \alpha_k g(n-k) + Gu(n). \quad (8.40)$$

Если теперь  $\tilde{R}(m)$  будет обозначать автокорреляционную функцию  $g(n)$ , то

$$\begin{aligned} \tilde{R}(m) = E[g(n)g(n-m)] &= \sum_{k=1}^p \alpha_k E[g(n-k)g(n-m)] + E[Gu(n)g(n-m)] \\ &\times (n-m) = \sum_{k=1}^p \alpha_k \tilde{R}(m-k), \quad m \neq 0, \end{aligned} \quad (8.41)$$

поскольку  $E[u(n)g(n-m)] = 0, m > 0$ , вследствие некоррелированности предшествующим  $u(n)$ . Для  $m=0$  получаем

$$\tilde{R}(0) = \sum_{k=1}^p \alpha_k \tilde{R}(k) + GE[u(n)g(n)] = \sum_{k=1}^p \alpha_k \tilde{R}(k) + G^2, \quad (8.42)$$

где  $E[u(n)g(n)] = E[u(n)(Gu(n) + \text{члены, предшествующие } n)] = G^2$ . Поскольку энергия отклика на  $Gu(n)$  должна совпадать с энергией сигнала, получаем

$$\tilde{R}(m) = R_n(m), \quad 0 \leq m \leq p \quad (8.43)$$

или

$$G^2 = R_n(0) - \sum_{k=1}^p \alpha_k R_n(k), \quad (8.44)$$

т. е. имеем то же самое, что и в случае импульсного возбуждения.

### 8.3. Решения уравнений линейного предсказания

Для эффективного использования метода линейного предсказания необходимо разработать эффективные алгоритмы решения системы линейных уравнений. Хотя можно использовать различные методы решения  $p$  уравнений с  $p$  неизвестными, все они оказываются различными по объему вычислений. Учитывая специальные свойства матрицы системы, решение можно получить значительно быстрее, чем в общем случае. В данном разделе подробно рассматриваются два метода получения параметров линейного предсказания и затем сравниваются и противопоставляются некоторые свойства полученных решений.

#### 8.3.1. Решения на основе разложения Холецкого для ковариационного метода [3]

Система уравнений, решаемая при ковариационном методе, имеет вид

$$\sum_{k=1}^p \alpha_k \varphi_n(i, k) = \varphi_n(i, 0), \quad i = 1, 2, \dots, p, \quad (8.45)$$

или в матричной форме

$$\Phi \alpha = \psi, \quad (8.46)$$

где  $\Phi$  — положительно определенная симметричная матрица из элементов  $\varphi_n(i, j)$ , а  $\alpha$  и  $\psi$  — вектор-столбцы с элементами  $\alpha_i$  и  $\varphi_n(i, 0)$  соответственно. Решение системы уравнений (8.45) можно получить с использованием эффективного алгоритма с учетом симметрии и положительной определенности матрицы  $\Phi$ . Получающийся при этом алгоритм называют методом Холецкого (или иногда методом квадратных корней) [3]. В этом методе матрица выражается в виде

$$\Phi = \mathbf{V} \mathbf{D} \mathbf{V}^t, \quad (8.47)$$

где  $\mathbf{V}$  — нижняя треугольная матрица (элементы главной диагонали которой — единицы), а  $\mathbf{D}$  — диагональная матрица. Верхний

индекс  $t$  означает транспонирование. Элементы матриц  $V$  и  $D$  легко определить из (8.47), приравняв  $(i, j)$ -элементы для фиксированных  $(i, j)$  слева и справа, что позволяет получить

$$\varphi_n(i, j) = \sum_{k=1}^j V_{ik} d_k V_{jk}, \quad 1 \leq j \leq i-1, \quad (8.48)$$

или

$$V_{ij} d_j = \varphi_n(i, j) - \sum_{k=1}^{j-1} V_{ik} d_k V_{jk}, \quad 1 \leq j \leq i-1, \quad (8.49)$$

и для диагональных элементов

$$\varphi_n(i, i) = \sum_{k=1}^i V_{ik} d_k V_{ik} \quad (8.50)$$

или

$$d_i = \varphi_n(i, i) - \sum_{k=1}^{i-1} V_{ik}^2 d_k, \quad i \geq 2, \quad (8.51)$$

если

$$d_1 = \varphi_n(1, 1). \quad (8.52)$$

Для иллюстрации использования соотношений (8.47)–(8.52) рассмотрим в качестве примера случай, когда  $p=4$  и  $\varphi_n(i, j) = \varphi_{ij}$ . Уравнения (8.47) в этом случае имеют вид

$$\begin{vmatrix} \varphi_{11} & \varphi_{21} & \varphi_{31} & \varphi_{41} \\ \varphi_{21} & \varphi_{22} & \varphi_{23} & \varphi_{41} \\ \varphi_{31} & \varphi_{32} & \varphi_{33} & \varphi_{43} \\ \varphi_{41} & \varphi_{42} & \varphi_{43} & \varphi_{44} \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 & 0 \\ V_{21} & 1 & 0 & 0 \\ V_{31} & V_{32} & 1 & 0 \\ V_{41} & V_{42} & V_{43} & 1 \end{vmatrix} \begin{vmatrix} d_1 & 0 & 0 & 0 \\ 0 & d_2 & 0 & 0 \\ 0 & 0 & d_3 & 0 \\ 0 & 0 & 0 & d_4 \end{vmatrix} \begin{vmatrix} 1 & V_{21} & V_{31} & V_{41} \\ 0 & 1 & V_{32} & V_{42} \\ 0 & 0 & 1 & V_{43} \\ 0 & 0 & 0 & 1 \end{vmatrix}.$$

Решение относительно  $d_1, d_4$  и  $V_{ij}$  начнем с (8.52) при  $i=1$ , что дает  $d_1 = \varphi_{11}$ . Используя (8.49) для  $i=2, 3, 4$ , получим  $V_{21}, V_{31}$  и  $V_{41}$  в виде  $V_{21}d_1 = \varphi_{21}$ ,  $V_{31}d_1 = \varphi_{31}$ ,  $V_{41}d_1 = \varphi_{41}$ ,  $V_{21} = \varphi_{21}/d_1$ ,  $V_{41} = \varphi_{41}/d_1$ . Используя (8.51) для  $i=2$ , получаем  $d_2 = \varphi_{22} - V_{21}^2 d_1$ . Из (8.49) для  $i=3$  и  $4$  имеем  $V_{32}d_2 = \varphi_{32} - V_{31}d_1 V_{21}$ ,  $V_{42}d_2 = \varphi_{42} - V_{41}d_1 V_{21}$  или  $V_{32} = (\varphi_{32} - V_{31}d_1 V_{21})/d_2$ ,  $V_{42} = (\varphi_{42} - V_{41}d_1 V_{21})/d_2$ . Из (8.51) при  $i=3$  получаем  $d_3$ , из (8.49) для  $i=4$  находим  $V_{43}$  и, наконец, из (8.51) при  $i=4$  определяем  $d_4$ . Получим матрицы  $V$  и  $D$ . Легко определить вектор-столбец  $\alpha$  с помощью двухшаговой процедуры. Из (8.46) и (8.47) находим

$$VDV^t \alpha = \psi, \quad (8.53)$$

что можно переписать в виде

$$VY = \psi \text{ и } DV^t \alpha = Y \quad (8.54); \quad (8.55)$$

или

$$V^t \alpha = D^{-1} Y. \quad (8.56)$$

Таким образом, при известной матрице  $V$  уравнения (8.54) можно разрешить относительно вектор-столбца  $Y$ , используя простую рекурсивную процедуру

$$Y_i = \psi_i - \sum_{j=1}^{i-1} V_{ij} Y_j, \quad p \geq i \geq 2, \quad (8.57)$$

с начальным условием

$$Y_1 = \psi_1. \quad (8.58)$$

Аналогично, имея решение  $Y$ , можно из (8.56) получать  $\alpha$  на основе рекурсивной процедуры вида

$$\alpha_i = Y_i/d_i - \sum_{j=i+1}^p V_{ji} \alpha_j, \quad 1 \leq i \leq p-1, \quad (8.59)$$

с начальными условиями

$$\alpha_p = Y_p/d_p. \quad (8.60)$$

Индекс  $i$  изменяется от  $i=p-1$  до  $i=1$  в порядке убывания.

Для иллюстрации использования алгоритма (8.57)–(8.60) продолжим рассмотрение примера. Сначала определим  $Y'_i$ , полагая, что  $V$  и  $D$  известны. В матричной форме имеем уравнения

$$\begin{vmatrix} 1 & 0 & 0 & 0 \\ V_{21} & 1 & 0 & 0 \\ V_{31} & V_{32} & 1 & 0 \\ V_{41} & V_{42} & V_{43} & 1 \end{vmatrix} \begin{vmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{vmatrix} = \begin{vmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \end{vmatrix}.$$

Из (8.57) и (8.58) получаем:  $Y_1 = \psi_1$ ,  $Y_2 = \psi_2 - V_{21}Y_1$ ,  $Y_3 = \psi_3 - V_{31}Y_1 - V_{32}Y_2$ ,  $Y_4 = \psi_4 - V_{41}Y_1 - V_{42}Y_2 - V_{43}Y_3$ . Используя  $Y'_i$ s, решим уравнения (8.56) вида

$$\begin{vmatrix} 1 & V_{21} & V_{31} & V_{41} \\ 0 & 1 & V_{32} & V_{42} \\ 0 & 0 & 1 & V_{43} \\ 0 & 0 & 0 & 1 \end{vmatrix} \begin{vmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{vmatrix} = \begin{vmatrix} 1/d_1 & 0 & 0 & 0 \\ 0 & 1/d_2 & 0 & 0 \\ 0 & 0 & 1/d_3 & 0 \\ 0 & 0 & 0 & 1/d_4 \end{vmatrix} \begin{vmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{vmatrix} = \begin{vmatrix} Y_1/d_1 \\ Y_2/d_2 \\ Y_3/d_3 \\ Y_4/d_4 \end{vmatrix}.$$

Из (8.59) и (8.60) получаем:  $\alpha_4 = Y_4/d_4$ ,  $\alpha_3 = Y_3/d_3 - V_{43}\alpha_4$ ,  $\alpha_2 = Y_2/d_2 - V_{32}\alpha_3 - V_{42}\alpha_4$ ,  $\alpha_1 = Y_1/d_1 - V_{21}\alpha_2 - V_{31}\alpha_3 - V_{41}\alpha_4$ , что и завершает решение ковариационных уравнений.

Используя разложение Холецкого, можно получить простое выражение для минимальной погрешности ковариационного метода через вектор  $Y$  и матрицу  $D$ . Напомним, что погрешность предсказания в ковариационном методе имеет вид

$$E_n = \varphi_n(0, 0) - \sum_{k=1}^p \alpha_k \varphi_n(0, k), \quad (8.61)$$

или в матричных обозначениях

$$E_n = \varphi_n(0, 0) - \alpha^t \psi. \quad (8.62)$$



Подставляя вместо  $\alpha^t$  его выражение из (8.56) в виде  $Y^t D^{-1} V^{-1}$ , получим

$$E_n = \varphi_n(0, 0) - Y^t D^{-1} V^{-1} \psi. \quad (8.63)$$

Используя (8.54), имеем

$$E_n = \varphi_n(0, 0) - Y^t D^{-1} Y, \quad (8.64)$$

или

$$E_n = \varphi_n(0, 0) - \sum_{k=1}^p Y_k^2 / d_k. \quad (8.65)$$

Таким образом, минимальное значение среднего квадрата ошибки можно непосредственно определить через вектор  $Y$  и матрицу  $D$ . Более того, (8.65) можно использовать для вычисления  $E_n$  при различных значениях  $p$  вплоть до значения, используемого при решении уравнений. Следовательно, можно проследить, как изменяется мощность погрешности предсказания при увеличении числа коэффициентов предсказания.

### 8.3.2. Алгоритм Дарбина для рекурсивного решения автокорреляционных уравнений [2]

Для автокорреляционного метода матричное уравнение относительно параметров предсказания имеет вид

$$\sum_{k=1}^p \alpha_k R_n(|i-k|) = R_n(i), \quad 1 \leq i \leq p, \quad (8.66)$$

так как матрица системы имеет тридиагональную форму, существует ряд специальных алгоритмов решения этой системы уравнений. Хотя наиболее известными и популярными являются методы Левинсона и Робинсона [1], эффективнее с точки зрения изложения является алгоритм Дарбина [2] (для простоты индекс у автокорреляционной функции опущен):

$$E^{(0)} = R(0); \quad (8.67)$$

$$k_i = [R(i) - \sum_{j=1}^{i-1} \alpha_j^{(i-1)} R(i-j)] / E^{(i-1)}, \quad 1 \leq i \leq p; \quad (8.68)$$

$$\alpha_i^{(i)} = k_i; \quad (8.69)$$

$$\alpha_j^{(i)} = \alpha_j^{(i-1)} - k_i \alpha_{i-j}^{(i-1)}, \quad 1 \leq j \leq i-1; \quad (8.70)$$

$$E^{(i)} = (1 - k_i^2) E^{(i-1)}. \quad (8.71)$$

Уравнения (8.67) — (8.71) решаются рекурсивно для  $i=1, 2, \dots, p$ , и окончательное решение принимает вид

$$\alpha_j = \alpha_j^{(p)}, \quad 1 \leq j \leq p. \quad (8.72)$$

Отметим, что при вычислении параметров предсказания для модели порядка  $p$  получаются и все параметры для предсказания меньших порядков, т. е.  $\alpha_j^{(i)}$  — это  $j$ -й параметр предсказателя порядка  $i$ .

Для иллюстрации применения рассмотренного выше алгоритма вычислим параметры предсказателя второго порядка. Исходные матричные уравнения имеют вид

$$\begin{vmatrix} R(0) & R(1) \\ R(1) & R(0) \end{vmatrix} \begin{vmatrix} \alpha_1 \\ \alpha_2 \end{vmatrix} = \begin{vmatrix} R(1) \\ R(2) \end{vmatrix}.$$

Используя приведенный алгоритм, имеем:

$$E^{(0)} = R(0); \quad k_1 = R(1)/R(0); \quad \alpha_1^{(1)} = R(1)/R(0);$$

$$E^{(1)} = \frac{R^2(0) - R^2(1)}{R(0)}; \quad k_2 = \frac{R(2)R(0) - R^2(1)}{R^2(0) - R^2(1)};$$

$$\alpha_2^{(2)} = \frac{R(2)R(0) - R^2(1)}{R^2(0) - R^2(1)}; \quad \alpha_1^{(2)} = \frac{R(1)R(0) - R(1)R(2)}{R^2(0) - R^2(1)};$$

$$\alpha_1 = \alpha_1^{(2)}; \quad \alpha_2 = \alpha_2^{(2)}.$$

Величины  $E^{(i)}$  в (8.71) представляют собой мощности погрешностей предсказания для предсказателя порядка  $i$ . Таким образом, на каждом шаге вычислений можно контролировать мощность погрешности. Коэффициенты корреляции можно заменить нормированными коэффициентами  $r(i)$ , что не изменит решения уравнения, но при этом величины  $E^{(i)}$  следует интерпретировать как нормализованную погрешность. Если обозначить ее через  $V^{(i)}$ , то

$$V^{(i)} = \frac{E^{(i)}}{R(0)} = 1 - \sum_{k=1}^i \alpha_k r(k), \quad (8.73)$$

где

$$0 < V^{(i)} \leq 1, \quad i \geq 0. \quad (8.74)$$

Можно показать, что нормированная погрешность при  $i=p$  может быть представлена в виде

$$V^{(p)} = \prod_{i=1}^p (1 - k_i^2), \quad (8.75)$$

где величины  $k_i$  удовлетворяют условию

$$-1 \leq k_i \leq 1. \quad (8.76)$$

Это ограничение параметров  $k_i$  весьма важно, ибо [1, 18] оно является необходимым и достаточным условием того, чтобы все корни полинома  $A(z)$  лежали внутри единичной окружности. Это гарантирует устойчивость системы  $H(z)$ . К сожалению, доказательство этого факта увело бы нас далеко в сторону, но его отсутствие в данной книге не снижает важности приведенного результата. Более того, можно показать, что в ковариационном методе нет таких условий устойчивости.

### 8.3.3. Постановка задачи и ее решение на основе лестничного фильтра [11]

Как было показано выше, оба метода вычисления параметров предсказания включают в себя два этапа: оценивание матрицы корреляций и решение системы линейных уравнений. Эти методы широко и успешно используются применительно к обработке речевых сигналов. Вместе с тем к настоящему времени развит другой класс методов, называемых методами на основе лестничного фильтра, в которых оба этапа в известном смысле объединены в один рекурсивный алгоритм оценивания параметров линейного предсказания. Для того чтобы проследить связь между этими методами, полезно начать с алгоритма Дарбина. Прежде всего отметим, что коэффициенты предсказания на  $i$ -й интерации являются параметрами предсказателя  $i$ -го порядка. Используя эти коэффициенты, можно определить

$$A^{(i)}(z) = 1 - \sum_{k=1}^i \alpha_k^{(i)} z^{-k} \quad (8.77)$$

как передаточную функцию обратного фильтра порядка  $i$  (или фильтр погрешности предсказания). Если на входе этого фильтра имеется сигнал  $s_n(m) = s(n+m)w(m)$ , то на его выходе погрешность предсказания  $e_n^{(i)}(m) = e^{(i)}(n+m)$ , где

$$e^{(i)}(m) = s(m) - \sum_{k=1}^i \alpha_k^{(i)} s(m-k). \quad (8.78)$$

Отметим, что далее для простоты опускается индекс  $n$ , который означает, что рассматривается сегмент сигнала, расположенный на  $n$ -м отсчете. Используя  $z$ -преобразование, запишем (8.78) в виде

$$E^{(i)}(z) = A^{(i)}(z) s(z). \quad (8.79)$$

Подставляя (8.70) в (8.77), получаем рекурсивную формулу для  $A^{(i)}(z)$  через  $A^{(i-1)}(z)$ , т. е.

$$A^{(i)}(z) = A^{(i-1)}(z) - k_i z^{-i} A^{(i-1)}(z^{-1}) \quad (8.80)$$

(см. задачу 8.5). Подставляя (8.80) в (8.79), получаем

$$E^{(i)}(z) = A^{(i-1)}(z) s(z) - k_i z^{-i} A^{(i-1)}(z^{-1}) s(z). \quad (8.81)$$

Первый член в (8.81), очевидно, является  $z$ -преобразованием погрешности предсказания для предсказателя  $(i-1)$ -го порядка. Второй член можно интерпретировать аналогичным образом, если ввести обозначения

$$B^{(i)}(z) = z^{-i} A^{(i)}(z^{-1}) s(z). \quad (8.82)$$

Обратное преобразование от  $B^{(i)}(z)$  есть

$$b^{(i)}(m) = s(m-i) - \sum_{k=1}^i \alpha_k^{(i)} s(m+k-i). \quad (8.83)$$

Это уравнение предполагает, что мы предсказываем  $s(m-i)$  по  $i$  отсчетам входного сигнала  $\{s(m-i+k), k=1, 2, \dots, i\}$ , которые следуют за  $s(m-i)$ . Таким образом, можно сказать, что  $b^{(i)}(m)$  — погрешность возвратного предсказания. На рис. 8.2 показано, что

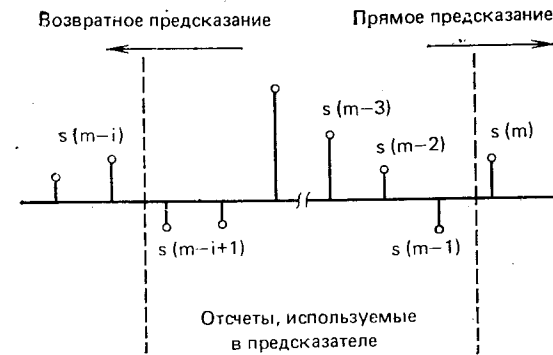


Рис. 8.2. Иллюстрация прямой и возвратной погрешностей для предсказателя  $i$ -го порядка

$i$  отсчетов, используемых для предсказания, в данном случае совпадают с отсчетами, которые применяются для предсказания текущего значения (8.78). Возвращаясь к (8.81), видим, что погрешность предсказания может быть представлена в виде

$$e^{(i)}(m) = e^{(i-1)}(m) - k_i b^{(i-1)}(m-1). \quad (8.84)$$

Подставляя (8.80) в (8.82), получим

$$B^{(i)}(z) = z^{-i} A^{(i-1)}(z^{-1}) s(z) - k_i A^{(i-1)}(z) s(z), \quad (8.85)$$

или

$$B^{(i)}(z) = z^{-1} B^{(i-1)}(z) - k_i E^{(i-1)}(z). \quad (8.86)$$

Таким образом,  $i$ -я возвратная погрешность

$$b^{(i)}(m) = b^{(i-1)}(m-1) - k_i e^{(i-1)}(m). \quad (8.87)$$

Уравнения (8.84) и (8.85) определяют погрешность прямого и возвратного предсказаний для предсказателя порядка  $i$  через соответствующие погрешности для предсказателя порядка  $(i-1)$ . Использование предсказателя нулевого порядка эквивалентно отсутствию предсказания вообще, т. е.

$$e^{(0)}(m) = b^{(0)}(m) = s(m), \quad (8.88)$$

Таким образом, уравнения (8.84) и (8.85) можно представить в виде структурной схемы (рис. 8.3). Такая схема называется лестничной. Очевидно, используя  $p$  каскадов лестничного фильтра, на выходе последнего из них можно получить погрешность предсказания, как это изображено на рис. 8.3. Таким образом, на рис. 8.3 представлена цифровая реализация фильтра погрешности предсказания с передаточной функцией  $A(z)$ .

Полученная схема является непосредственным следствием алгоритма Дарбина, и параметры  $k_i$  можно вычислять с использованием уравнений (8.67)–(8.72). Параметры предсказания как таковые на рис. 8.3 отсутствуют. Итакура [4, 6] показал, что  $k_i$

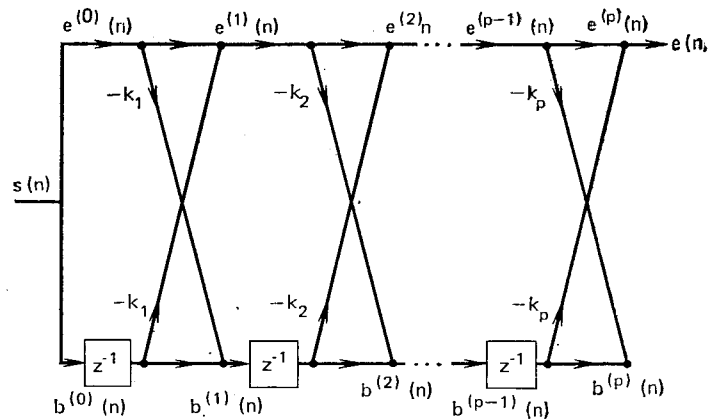


Рис. 8.3. Структурная схема реализации лестничного фильтра

можно получить непосредственно по погрешностям предсказания и в силу особенностей лестничной схемы совокупность коэффициентов  $k_i$ ,  $i=1, 2, \dots, p$ , можно получить без вычисления коэффициентов предсказания в соответствии с соотношением

$$k_i = \frac{\sum_{m=0}^{N-1} e^{(i-1)}(m) b^{(i-1)}(m-1)}{\left\{ \sum_{m=0}^{N-1} (e^{(i-1)}(m))^2 \sum_{m=0}^{N-1} (b^{(i-1)}(m-1))^2 \right\}^{1/2}}. \quad (8.89)$$

Это выражение — разновидность нормированной взаимно-корреляционной функции, т. е. показатель корреляции между погрешностью предсказания и возвратной погрешностью. Поэтому параметры  $k_i$  называют частными корреляциями [4, 6]. Уравнение (8.89) эквивалентно (8.68) после подстановки (8.78) и (8.83) в (8.89).

Заменяя (8.68) на (8.89) в алгоритме Дарбина, можно, как и ранее, получить параметры предсказания. Таким образом, анализ на основе частных корреляций приводит к несколько иному подходу, чем при обращении матриц, и дает результаты, совпадающие с автокорреляционным методом, т. е. совокупность параметров частных корреляций эквивалентна совокупности параметров предсказания, минимизирующих средний квадрат погрешности предсказания. Этот метод открывает новый класс методов, основанных на лестничном фильтре (рис. 8.3) [11].

Аналогично Бург [12] разработал алгоритм, основанный на минимизации суммы среднего квадрата прямой и возвратной погрешностей на рис. 8.3, т. е.

$$\tilde{E}^{(i)} = \sum_{m=0}^{N-1} [(e^{(i)}(m))^2 + (b^{(i)}(m))^2]. \quad (8.90)$$

Подставляя (8.84) и (8.87) в (8.90) и дифференцируя  $\tilde{E}^{(i)}$  по  $k_i$ , получаем

$$\frac{\partial \tilde{E}^{(i)}}{\partial k_i} = -2 \sum_{m=0}^{N-1} [e^{(i-1)}(m) - k_i b^{(i-1)}(m-1)] b^{(i-1)}(m-1) - 2 \sum_{m=0}^{N-1} [b^{(i-1)}(m-1) - k_i e^{(i-1)}(m)] e^{(i-1)}(m). \quad (8.91)$$

Приравнявая производную нулю, находим

$$k_i = \frac{2 \sum_{m=0}^{N-1} [e^{(i-1)}(m) b^{(i-1)}(m-1)]}{\sum_{m=0}^{N-1} [e^{(i-1)}(m)]^2 + \sum_{m=0}^{N-1} [b^{(i-1)}(m-1)]^2}. \quad (8.92)$$

Можно показать, что оценки  $k_i$  на основе (8.92) удовлетворяют соотношению

$$-1 \leq k_i \leq 1. \quad (8.93)$$

Следует, однако, иметь в виду, что оценки  $k_i$  в соответствии с (8.92) отличаются от оценок (8.89) или, что то же самое, от автокорреляционного метода.

Отметим в заключение, что вычисление коэффициентов предсказания и параметров  $k$  включает в себя следующие шаги:

1. Начальные условия  $e^{(0)}(m) = s(m) = b^{(0)}(m)$ .
2. Вычислить  $k_1 = \alpha_1^{(1)}$  из (8.92).
3. Определить прямую и возвратную погрешности предсказания  $e^{(1)}(m)$  и  $b^{(1)}(m)$  по (8.84) и (8.87).
4. Установить  $i=2$ .
5. Определить  $k_i = \alpha_i^{(i)}$  из (8.92).
6. Определить  $\alpha_j^{(i)}$  для  $j=1, 2, \dots, i-1$  из (8.70).
7. Определить  $e^{(i)}(m)$  и  $b^{(i)}(m)$  из (8.84) и (8.87).
8. Установить  $i=i+1$ .
9. Если  $i$  меньше или равно  $p$ , идти к 5.
10. Алгоритм закончен.

Между изложенным методом, а также автокорреляционным и ковариационным методами существует ряд различий. Основное из них состоит в том, что в лестничном методе коэффициенты предсказания оцениваются непосредственно по речевому сигналу без промежуточного вычисления автокорреляционной функции.

Кроме того, метод гарантирует получение устойчивого фильтра без использования окон. По этим причинам подход на основе лестничного фильтра является важным и предпочтительным способом реализации линейного предсказания.

#### 8.4. Сравнение методов решения уравнений линейного предсказания

Выше уже обсуждались различия между ковариационным, корреляционным подходами и методом на основе лестничного фильтра. Здесь рассматриваются вопросы практического использования полученных уравнений. Эти вопросы включают в себя вычислительную сторону задачи, цифровую и физическую устойчивость получающихся решений и вопросы выбора количества полюсов или каскадов модели. Начнем с вычислительных аспектов задачи, связанных с получением параметров предсказания по речевому сигналу.

Два основных вопроса при вычислении коэффициентов предсказания состоят в объеме памяти и количестве умножения. В табл. 8.1 (по результатам Портнова и др. [13] и Макхоула

Таблица 8.1

Объем вычислений при решении уравнений линейного предсказания

	Ковариационный метод	Автокорреляционный метод	Лестничный метод
	Разложение Холецкого	Метод Дарбина	Метод Бурга
Память: данные матрица	$N_1$	$N_2$	$3N_3$
окно	Пропорционально $p^2/2$	Пропорционально $p$	—
Вычисления (умножения):	0	$N_2$	—
взвешивание корреляция	Пропорционально $N_1 p$	Пропорционально $N_2 p$	—
матричное решение	Пропорционально $p^3$	Пропорционально $p^2$	$5N_3 p$

[11]) содержится требуемый объем вычислений для трех рассмотренных методов. С точки зрения памяти, ковариационный метод требует наличия  $N_1$  ячеек для данных и  $p^2/2$  ячеек для ковариационной матрицы. Для автокорреляционного метода требуется  $N_2$  ячеек памяти как для данных, так и для окна из  $p$  ячеек для хранения автокорреляционной матрицы. Для лестничного метода необходимо наличие  $3N_3$  ячеек памяти для исходных данных и двух погрешностей предсказания. Для того чтобы подчеркнуть

различие в объеме исходных данных, в случае ковариационного метода этот объем обозначен через  $N_1$ , автокорреляционного —  $N_2$ , лестничного —  $N_3$ . Вопрос выбора этого объема будет обсуждаться далее. А сейчас, полагая, что  $N_1$ ,  $N_2$  и  $N_3$  сравнимы, получаем, что автокорреляционный и ковариационный методы требуют меньше памяти, чем лестничный метод.

Количество умножений, необходимое для реализации каждого метода, показано в нижней части табл. 8.1. Для ковариационного метода вычисление матрицы требует около  $N_1 p$  умножений, а решение матричного уравнения (с использованием разложения Холецкого) осуществляется с использованием  $p^3$  умножений [Портнов и другие дали точное значение:  $(p^3 + 9p^2 + 2p)/6$  операций умножения,  $p$  — деления и  $p$  — извлечения квадратного корня]. Для автокорреляционного метода вычисление корреляционной матрицы требует проведения  $N_2 p$  умножений, а решение автокорреляционных уравнений — около  $p^2$  умножений. Таким образом, если  $N_1 \approx N_2$  и  $N_1 \gg p$ ,  $N_2 \gg p$ , то автокорреляционный метод требует меньше вычислений, чем ковариационный. Однако в большинстве случаев при обработке речевых сигналов количество операций умножения для вычисления автокорреляционной функции значительно превосходит объем вычислений при решении уравнения, поэтому время вычислений в обоих случаях примерно одинаково. Для вычисления частных корреляций в лестничном методе необходимо осуществить  $5N_3 p$  умножений<sup>1</sup>. Таким образом с точки зрения эффективности вычислительной процедуры лестничный метод наименее эффективен. Однако при решении вопроса о его использовании следует иметь в виду ряд преимуществ этого метода по сравнению с другими.

Другой аспект сопоставления различных подходов связан с устойчивостью полученной системы:

$$H(z) = G/A(z). \quad (8.94)$$

Этот фильтр устойчив, если все его полюсы лежат строго внутри единичной окружности. Полюсы фильтра  $H(z)$  совпадают с нулями полинома знаменателя  $A(z)$ , т. е.

$$A(z) = 1 - \sum_{k=1}^p \alpha_k z^{-k}. \quad (8.95)$$

Как утверждалось выше, для автокорреляционного метода все корни лежат внутри единичной окружности, т. е. устойчивость  $H(z)$  гарантирована. Следует отметить, что эти теоретические гарантии могут нарушаться на практике, если автокорреляционная функция вычисляется с недостаточной точностью. Встречающиеся в этом случае округления при вычислении автокорреляционной функции могут привести к плохой обусловленности автокорреляционной матрицы. Маркел и Грей показали, что этот эф-

<sup>1</sup> Макхоул [11] предложил метод вычисления частных корреляций с той же эффективностью, что и при ковариационном методе.

фект можно устранить применением предсказаний, приводящих к выравниванию спектра [1]. Использование предсказаний позволяет уменьшить разрядность вычислителя при сохранении устойчивости предсказателя. Алгоритм Дарбина позволяет легко проверить устойчивость, поскольку параметры  $k_i$  (частные корреляции) должны удовлетворять условию

$$-1 \leq k_i \leq 1. \quad (8.96)$$

Таким образом, если при вычислении коэффициентов предсказания любая из величин  $k_i$  не удовлетворяет (8.96), то, как известно, корни лежат вне единичной окружности.

Для ковариационного метода невозможно гарантировать устойчивость фильтра-предсказателя. Однако на практике при достаточно большом числе отсчетов на интервале оценивания получаемый предсказатель почти всегда устойчив. Это объясняется тем, что при большом числе отсчетов на интервале оценивания ковариационный и автокорреляционный методы дают почти одинаковый результат.

Для лестничного метода устойчивость гарантирована, ибо коэффициенты предсказания получаются по коэффициентам частных корреляций, которые по определению удовлетворяют (8.96). Кроме того, устойчивость сохраняется и при вычислениях с использованием вычислителя с конечной разрядной сеткой [1]. При возникновении сомнений в устойчивости предсказателя необходимо определить корни полинома предсказателя. Если обнаружено, что корни расположены вне единичной окружности, то простая процедура, состоящая в отражении корней внутрь окружности, позволяет получить устойчивый полином с той же частотной характеристикой, что и неустойчивый.

Два других аспекта проблемы решения уравнений линейного предсказания сводятся к выбору порядка предсказателя  $p$  и длины интервала анализа  $N$ . Выбор  $p$  определяется частотой дискретизации и не зависит от используемого метода. Поскольку подлежащий анализу речевой спектр характеризуется в общем случае вкладом голосового тракта со средней плотностью примерно два полюса (или один комплексный полюс) на каждый килогерц, то для представления полного вклада в речевой спектр необходимо наличие  $F_s$  полюсов, где  $F_s$  — частота дискретизации в килогерцах. Таким образом, при дискретизации с частотой 10 кГц необходимо десять полюсов для представления голосового тракта. Кроме того, для описания источника возбуждения и излучения губами необходимо еще три-четыре полюса. Следовательно, общее число полюсов при частоте дискретизации 10 кГц составляет 13—14. Для подтверждения этого факта на рис. 8.4 представлена зависимость нормированной погрешности предсказания от порядка предсказателя  $p$  для вокализованной и невокализованной речей при частоте дискретизации 10 кГц. Хотя при возрастании  $p$  погрешность уменьшается, можно заметить, что уже при  $p \approx 13-14$  погрешность изменяется незначительно. Интересно, что

для невокализованного сигнала погрешность значительно больше, чем для вокализованного. Это, конечно, не является чем-то неожиданным, так как модель для невокализованного сигнала менее точна, чем для вокализованного. Дополнительное экспериментальное исследование поведения погрешности предсказания при различных  $p$  дано в последующих параграфах.

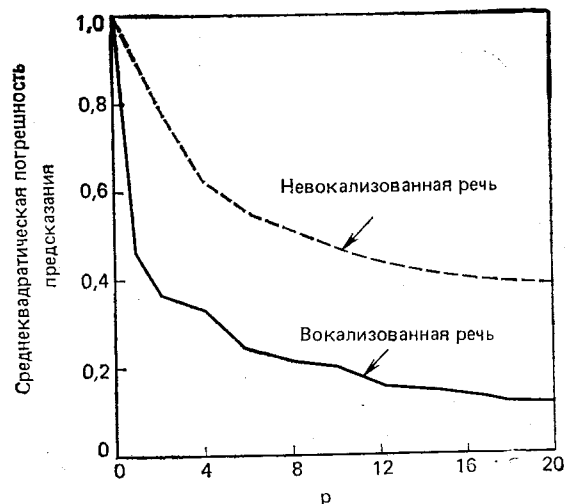


Рис. 8.4. Среднеквадратическая погрешность предсказания в зависимости от числа коэффициентов  $p$  [3]

Выбор длины интервала анализа  $N$  является чрезвычайно важным при использовании линейного предсказания. Выбирать следует как можно меньше, поскольку полный объем вычислений непосредственно связан с  $N$ . Для автокорреляционного метода показано, что для достижения хороших результатов значения  $N$  должны составлять несколько периодов основного тона [1, 2]. Поскольку при использовании автокорреляционного метода применяется временное окно, то для того чтобы эффект спадания окна на границах интервала не влиял на результаты, этот интервал должен быть достаточно большим. Так, в большинстве систем анализа на основе линейного предсказания используются окна от 100 до 400 отсчетов (при частоте дискретизации 10 кГц), причем предпочтение отдается более длинным окнам. Как для ковариационного, так и для лестничного методов выбор протяженности окна зависит от ряда условий. Поскольку взвешивание данных в этих случаях не требуется, нет и ограничений снизу на протяженность окна. Если анализ можно ограничить интервалом между импульсами основного тона (синхронно с ним), то достаточно, чтобы значение  $N$  составляло около  $2p$ . Но при использовании столь короткого интервала можно получить неудовлетворитель-