

## Лекция 3,4

### ВЕКТОРНОЕ КВАНТОВАНИЕ. СРАВНЕНИЕ ХАРАКТЕРИСТИК КВАНТОВАТЕЛЕЙ.

#### ВЕКТОРНОЕ КВАНТОВАНИЕ

Прежде чем перейти к описанию векторного квантования рассмотрим как оптимальное неравномерное скалярное квантование может быть реализовано при неизвестной функции плотности вероятностей квантуемой случайной величины  $X$ .

Пусть  $x_1, x_2, \dots, x_k$  последовательность значений случайной величины  $X$ , появившихся на входе квантователя за  $k$  последовательных моментов времени (реализация  $X$ ). Требуемая скорость квантования равна  $R_0$ . Для этого случая решение может быть найдено с помощью следующей рекуррентной процедуры (процедуры Макса-Ллойда):

1. Определяем число квантов по формуле  $M = 2^{R_0}$ .
2. Строим шкалу для равномерного квантования (вычисляем границы квантов и аппроксимирующие значения  $y_i$ ).
3. Выполняем скалярное квантование последовательности  $x_1, x_2, \dots, x_k$ .
4. Вычисляем новые аппроксимирующие значения  $y_i$ ,  $i = 1, \dots, M$  как среднее арифметическое значений случайной величины, попавших в  $i$ -й квант.
5. Вычисляем новые границы квантов по формуле  $x_i = (y_i + y_{i+1})/2$ .
6. Если условие остановки не выполнено, то возвращаемся к шагу 3.

В качестве условия остановки могут быть использованы следующие условия:

- Границы квантов, вычисленные на очередном шаге, отличаются от границ квантов, вычисленных на предыдущем шаге меньше, чем на заданную величину
- Число итераций превысило допустимое заданное число
- Среднеквадратическая ошибка квантования оказалась меньше заданной и.т.д.

Таким образом, мы получаем границы квантов и аппроксимирующие значения, используя которые можем выполнять скалярное неравномерное квантование случайной величины  $X$ .

Векторное квантование предполагает, что на вход квантователя поступает случайный вектор  $\mathbf{x} = (x_1, \dots, x_n)$ . Этот вектор заменяется одним (ближайшим в смысле выбранного критерия качества) вектором  $\mathbf{y} = (y_1, \dots, y_n)$  из аппроксимирующего множества  $Y = \{\mathbf{y}_1, \dots, \mathbf{y}_M\}$ . Аппроксимирующее множество  $Y$  часто называют *кодовой книгой*. Скорость кодирования равна  $R = \log_2 M / n$  бит/отсчет.

Для построения кодовой книги при неизвестной функции плотности вероятностей входного случайного вектора  $\mathbf{x}$  используется так называемый алгоритм Линде-Бузо-Грея. Этот алгоритм представляет собой обобщение процедуры Ллойда-Макса.

Пусть задано множество реализаций случайного вектора  $\mathbf{x}$ :

$\mathbf{x}_1 = (x_{11}, \dots, x_{1n})$ ,  $\mathbf{x}_2 = (x_{21}, \dots, x_{2n})$ , ...,  $\mathbf{x}_k = (x_{k1}, \dots, x_{kn})$ , пусть размер кодовой книги, которую необходимо построить  $M \ll k$ , ( $M = 2^{R_0 n}$ , где  $R_0$  - требуемая скорость кодирования). Процедура Линде-Бузо-Грея включает следующие шаги:

1. Строим начальную кодовую книгу, выбирая в качестве аппроксимирующих векторов  $\mathbf{y}_i$ ,  $i = 1, \dots, M$  произвольные  $M$  векторов  $\mathbf{x}_j$ ,  $j = 1, \dots, k$ .
2. Полагаем  $S_i = 0$ ,  $num_i = 0$ ,  $i = 1, \dots, M$ . Для всех векторов  $\mathbf{x}_j$ ,  $j = 1, \dots, k$

- вычисляем ошибку по формуле  $e_{opt} = \min_{i=1,...,M} (\mathbf{x}_j - \mathbf{y}_i)(\mathbf{x}_j - \mathbf{y}_i)^T$ , где  $T$  обозначает операцию транспонирования.
- вычисляем  $i_{opt} = \arg(\min_{i=1,...,M} (\mathbf{x}_j - \mathbf{y}_i)(\mathbf{x}_j - \mathbf{y}_i)^T)$ , т.е. индекс  $i$  при котором обеспечивается минимум ошибки для  $\mathbf{x}_j$ .

- вычисляем покомпонентные суммы векторов

$$S_{i_{opt}} = S_{i_{opt}} + \mathbf{x}_j.$$

- вычисляем число векторов  $\mathbf{x}_j$ , ближайших к каждому из аппроксимирующих значений  $\mathbf{y}_{i_{opt}}$

$$num_{i_{opt}} = num_{i_{opt}} + 1.$$

3. Для всех  $i = 1, ..., M$  вычисляем новые аппроксимирующие значения по формуле

$$\mathbf{y}_i = S_i / num_i$$

4. Если условие остановки не выполнено, то возвращаемся к шагу 2.

В качестве условия остановки могут быть использованы следующие:

- Аппроксимирующие значения, полученные на текущем шаге отличаются от аппроксимирующих значений, полученных на предыдущем шаге менее, чем на заданную величину
- Число итераций превысило допустимое заданное число
- Среднеквадратическая ошибка квантования оказалась меньше заданной.

Результатом применения процедуры Линде-Бузо-Грея является книга аппроксимирующих значений для вектора  $\mathbf{x}$ . При этом по аналогии с неравномерным оптимальным скалярным квантованием в процессе выполнения процедуры Линде-Бузо-Грея  $n$ -мерные аппроксимирующие значения смещаются в высоковероятные области под  $n$ -мерной функцией плотности вероятностей входного вектора.

Собственно процедура квантования состоит в выборе для входного вектора  $\mathbf{x}$  одного из векторов  $\mathbf{y}_i$ ,  $i = 1, ..., M$  по минимуму среднеквадратической ошибки  $\varepsilon = \min_{i=1,...,M} (\mathbf{x} - \mathbf{y}_i)(\mathbf{x} - \mathbf{y}_i)^T$ . Основным недостатком метода векторного квантования состоит в его высокой (переборной  $2^{nR_0}$ ) сложности. Например, при  $R_0 = 0.5$  бит/отсчет и  $n = 32$  размер кодовой книги составляет  $M = 2^{16}$ . Уменьшение сложности этой процедуры возможно за счет использования кодовых книг заданной структуры (например, решетчатых кодов). В этом случае для поиска ближайшего к входному вектору в кодовой книге может, например, быть использован алгоритм декодирования решетчатых кодов, известный как алгоритм Витерби.

## СРАВНЕНИЕ ХАРАКТЕРИСТИК КВАНТОВАТЕЛЕЙ

Основными характеристиками любого метода квантования являются ошибка и скорость квантования (кодирования). Рассмотрим вначале методы квантования с фиксированной скоростью, а затем обсудим как изменяются их характеристики, если выходные значения квантователя могут быть закодированы неравномерным кодом.

При различных методах квантования (кодирования непрерывных источников) могут быть получены различные функции  $R(\varepsilon)$  - скорости от ошибки. Наша задача – нахождение наилучшего кодирования, при котором кривая  $R(\varepsilon)$  проходит ниже других. Однако, указать наилучший метод квантования бывает довольно трудно и потому, что этот наилучший метод сложен, и также потому, что в некоторых случаях он просто неизвестен. Вместе с тем, в ряде случаев возможно найти наилучшую функцию скорости от ошибки не находя наилучшего кода. Эта функция называется *эпсилон-энтропией* и

обозначается как  $H(\varepsilon)$ . В теории кодирования непрерывных источников она играет ту же роль, что понятие энтропии в области кодирования дискретных источников. Однако, проблема состоит в том, что даже в случае скалярного квантования (аналог побуквенного кодирования дискретного источника) вычислить эту функцию удастся только для некоторых классов источников.

В частности, при использовании квадратичного критерия качества для *гауссовского источника без памяти*, т.е. источника, на выходе которого появляются независимые гауссовские случайные величины  $\mathbf{X}$ , энтальпия имеет вид

$$H(\varepsilon) = \begin{cases} \frac{1}{2} \log_2(\sigma^2 / \varepsilon), & \varepsilon \leq \sigma^2 \\ 0, & \varepsilon > \sigma^2 \end{cases}, \quad (2.1)$$

где  $\mathbf{X}$  имеет нулевое математическое ожидание и дисперсию  $\sigma^2$ .

В общем случае энтальпия  $H(\varepsilon)$  обладает следующими свойствами :

- Она является невозрастающей и выпуклой вниз функцией  $\varepsilon$
- Существует некоторое  $\varepsilon_0$  такое, что  $H(\varepsilon) = 0$  при всех  $\varepsilon \geq \varepsilon_0$ .

График  $H(\varepsilon)$  для источника независимых гауссовских случайных величин с нулевым средним и дисперсией  $\sigma^2 = 1$  приведен на Рис. 2.1. Очевидно, что в случае квантователя с фиксированной скоростью оптимальный неравномерный квантователь всегда имеет лучшую характеристику  $R(\varepsilon)$ , чем равномерный скалярный квантователь. На Рис.2.1 показана кривая  $R(\varepsilon)$  для того же источника. Как следует из Рис. 2.1 при использовании скалярного квантователя с фиксированной скоростью невозможно обеспечить скорость квантования менее 1 бита/отсчет. Отметим, что  $H(\varepsilon)$  для гауссовского источника, задаваемая формулой (2.1) является границей сверху для энтальпии любого источника без памяти с фиксированной дисперсией при использовании квадратичного критерия качества.

Переход от скалярного квантования к векторному позволяет, как и в случае перехода от побуквенного кодирования к блоковому кодированию дискретных источников, уменьшить избыточность кодирования, т.е.  $\max_{\varepsilon} (R(\varepsilon) - H(\varepsilon))$ . Можно показать, что для гауссовского источника без памяти при  $n \geq 2$  кривая  $R(\varepsilon)$  для векторного квантователя проходит ниже одноименной кривой для неравномерного скалярного квантователя. Кроме того векторный квантователь позволяет достичь скоростей кодирования меньших 1 бит/отсчет. На Рис. 2.1 приведены кривые  $R(\varepsilon)$  для гауссовского источника без памяти при использовании векторного квантования ( $n = 2, n = 6$ ).

Рассмотрим теперь источник, который порождает *стационарный гауссовский случайный процесс*. Это означает, что для любого  $n = 1, 2, \dots$  случайный вектор  $\mathbf{X} = (X_1, \dots, X_n)$  имеет гауссовское распределение вероятностей с ковариационной матрицей  $\Lambda_n$

$$f_n(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} |\Lambda_n|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \mathbf{m}) \Lambda_n^{-1} (\mathbf{x} - \mathbf{m})^T \right\},$$

а из условия стационарности следует, что каждый элемент  $\lambda_{ij}$  ковариационной матрицы  $\Lambda_n$  зависит только от абсолютного значения разности  $|i - j|$ , т.е.

$$\lambda_{ij} = \lambda_{ji} = \lambda_{|i-j|} \quad (2.2).$$

Матрицы, элементы которых удовлетворяют (2.2), называют *теплицевыми*. Пусть  $\tau = i - j$ ,  $\tau = -(n-1), \dots, 0, 1, \dots, n-1$ . Из (2.2) следует, что  $\lambda_{\tau} = \lambda_{-\tau}$  и что корреляционная матрица отрезка стационарного процесса полностью определяется  $n$  числами  $\lambda_{\tau}$ ,

$\tau = 0, 1, \dots, n-1$ , являющимися ковариационными моментами случайных величин  $X_i, X_j$ , отстоящими друг от друга на  $\tau$  единиц времени.

Будем полагать, что  $\mathbf{m} = \mathbf{0}$  и, что  $\lim_{\tau \rightarrow \infty} \lambda_\tau = 0$ . Рассмотрим ряд Фурье

$$\sum_{\tau=-\infty}^{\infty} \lambda_\tau e^{-j2\pi f\tau} = N(f), \quad -1/2 \leq f \leq 1/2.$$

Функция  $N(f)$  называется *спектральной плотностью мощности* процесса, порожденного стационарным источником. Если  $N(f)$  — непрерывная и ограниченная функция (ограничимся рассмотрением этого случая), то существует обратное преобразование, позволяющее по спектральной плотности мощности определить корреляционные моменты  $\lambda_\tau, \tau = 0, 1, \dots$ .

$$\lambda_\tau = \int_{-1/2}^{1/2} N(f) e^{j2\pi f\tau} df.$$

Основные свойства спектральной плотности мощности:

- $N(f)$  -действительная функция
- $N(f) = 2 \sum_{\tau=1}^{\infty} \lambda_\tau \cos 2\pi f\tau + \lambda_0$
- $N(f) = N(-f)$  и поэтому  $\lambda_\tau = \int_{-1/2}^{1/2} N(f) \cos 2\pi f\tau df$ .
- $N(f) \geq 0$  для всех  $f \in [-1/2, 1/2]$
- $\lambda_0 = \int_{-1/2}^{1/2} N(f) df$  -дисперсия случайного процесса.

Спектральная плотность мощности характеризует распределение мощности процесса по частотам.

Для гауссовского стационарного случайного процесса с дискретным временем и спектральной плотностью мощности  $N(f)$ , которая ограничена и интегрируема, энтрон-энтропия  $H(\varepsilon)$  при квадратичном критерии качества вычисляется по формулам

$$H(\varepsilon) = \frac{1}{2} \int_{-1/2}^{1/2} \log \left\{ \max \left\{ 1, \frac{N(f)}{\theta} \right\} \right\} df,$$

$$\int_{-1/2}^{1/2} \min \{ \theta, N(f) \} df = \varepsilon. \quad (2.3)$$

Если отсчеты стационарного гауссовского случайного процесса независимы и имеют одинаковую дисперсию  $\sigma^2$ , то (2.3) сводится к (2.1). Другими словами кривая  $H(\varepsilon)$  совпадает с приведенной на Рис.2.1. Если же отсчеты зависимы, то она будет лежать ниже кривой, приведенной на Рис.2.1.

Энтрон-энтропия стационарного случайного процесса в теории кодирования непрерывных источников играет ту же роль, что энтропия на сообщение дискретного стационарного источника в теории кодирования дискретных источников. Приведем без доказательства прямую и обратную теоремы кодирования для стационарного случайного процесса.

*Теорема 1 (прямая).* Для непрерывного по множеству значений и дискретного по времени стационарного эргодического источника с энтрон-энтропией  $H(\varepsilon)$  существует  $n_0$  такое, что для всех  $n > n_0$  и для любых  $\delta_1 > 0$  и  $\delta_2 > 0$  существует кодирование источника блоками длины  $n$ , для которого скорость кодирования  $R \leq H(\varepsilon) + \delta_1$  и ошибка

кодирования не превышает  $\varepsilon + \delta_2$ . Другими словами, при ошибке кодирования сколь угодно близкой к  $\varepsilon$  существует код со скоростью сколь угодно близкой к  $H(\varepsilon)$ .

*Теорема 2 (обратная).* Для источника, указанного в теореме 1, не существует кодирования, для которого одновременно ошибка была бы равной  $\varepsilon$  и скорость  $R < H(\varepsilon)$ .

Использование процедуры векторного квантования для гауссовского источника с зависимыми отсчетами позволяет при  $n \rightarrow \infty$  получить кривую  $R(\varepsilon)$ , лежащую ниже кривой эпсилон-энтропии на Рис. 2.1

Если требование фиксированной длины кодового слова на выходе квантователя не является обязательным, то альтернативой рассмотренным методам квантования может служить сочетание одного из видов квантования и неравномерного кодирования дискретных источников. Номера квантов на выходе квантователя рассматриваются как сообщения некоторого дискретного источника. Эти сообщения кодируются затем каким-либо неравномерным кодом (кодом Шеннона, кодом Хаффмена, арифметическим кодом и т.д.). При использовании такой двухступенчатой процедуры кодирования соотношение между рассмотренными выше типами квантователей меняется.

Сравним, например, равномерное и неравномерное скалярное квантование. Если кодировать неравномерным кодом (например кодом Хаффмена) дискретные сообщения на выходе каждого из квантователей, то средняя длина кодовых слов будет определяться энтропией квантованных сообщений, а ошибка квантования будет такой же, как при квантовании, поскольку кодирование неравномерными кодами с однозначным декодированием позволяет точно восстановить последовательность номеров квантов и поэтому дополнительной ошибки не вносит. Энтропия же квантованных сообщений определяется вероятностями попадания случайной величины в различные кванты. Она будет большей тогда, когда вероятности попадания в различные кванты имеют близкие значения и меньшей, когда они существенно различаются.

При оптимальном неравномерном квантовании кванты имеют малые длины в области высоковероятных значений квантуемой величины и большие длины в области маловероятных значений. Это означает, что вероятности попадания случайной величины в различные кванты при таком квантовании выравниваются и, следовательно, энтропия возрастает. При равномерном квантовании кванты имеют одинаковые длины, и вероятности попадания в них сильно различаются, т.е. энтропия должна быть в этом случае относительно малой. Таким образом, можно ожидать, что сочетание равномерного квантования и кодирования неравномерными кодами приведет к более сильному понижению кривой  $R(\varepsilon)$  по сравнению со случаем оптимального неравномерного квантования и кодирования теми же кодами. На Рис.2.1 приведена кривая  $R(\varepsilon)$  для случая квантования гауссовского источника без памяти с помощью двухступенчатой процедуры: неравномерное оптимальное скалярное квантование и последующее неравномерное кодирование. Как следует из рисунка эта кривая проходит ниже кривой, соответствующей векторному квантованию того же источника с  $n = 6$ .

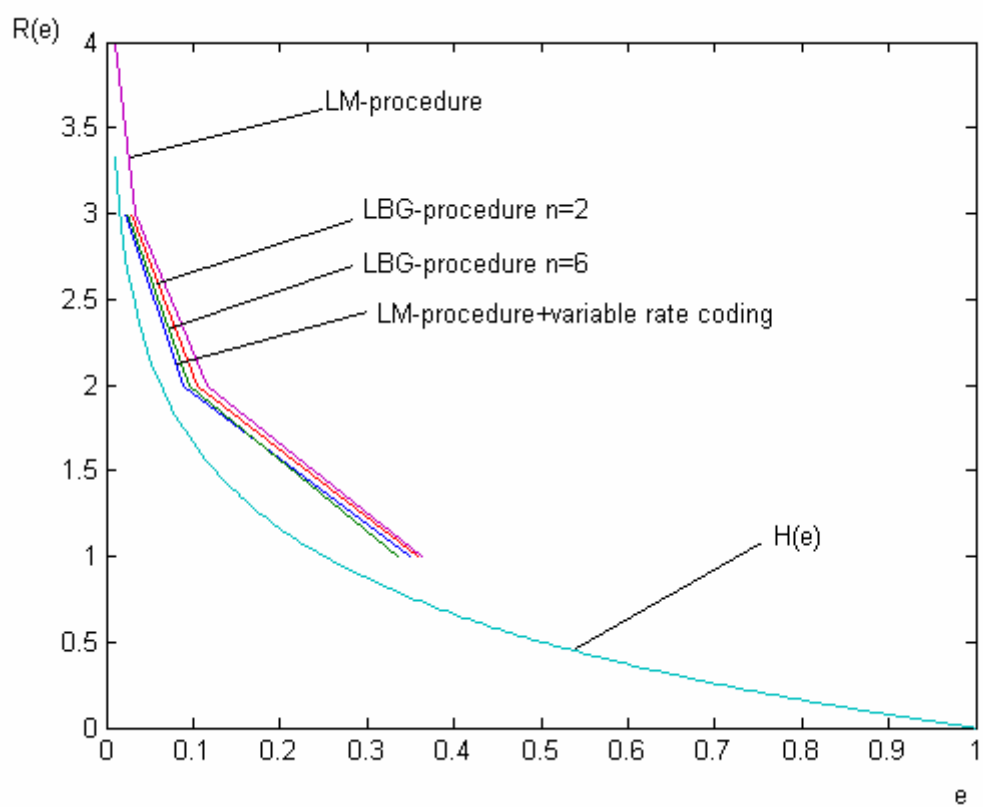


Рис. 2.1 Зависимость скорости кодирования от ошибки для гауссовского источника без памяти