Алгоритм оптимизации шкал квантования по произвольному показателю качества

Н.И. Глумов[1,2], М.В. Гашников[1,2]

[1]Самарский национальный исследовательский университет им. академика С.П. Королева, Московское шоссе 34А, Самара, Россия, 443086
[2]Институт систем обработки изображений РАН - филиал ФНИЦ «Кристаллография и фотоника» РАН, Молодогвардейская 151, Самара, Россия, 443001

Аннотация. В работе рассматривается задача построения шкал квантования, оптимальных по заданному критерию и удовлетворяющих выбранному ограничению. Рассматривается формальная постановка такой задачи оптимизации. Предлагается алгоритм построения квазиоптимальных шкал квантования, аппроксимирующих оптимальные шкалы с заданной точностью при соблюдении ограничения. Формулируются требования к критерию оптимизации и ограничению, обеспечивающие работоспособность алгоритма. Проводятся вычислительные эксперименты по построению квазиоптимальных шкал. Результаты экспериментов подтверждают преимущество построенных шкал над известными.

1. Введение

Под квантованием [1] обычно понимается разбиение диапазона некоторой величины на конечное число уровней квантования и округление значений этой величины до одного из ближайших к нему уровней. Другими словами, квантование это «загрубление» до заранее заданных уровней.

Квантование используется при решении различных задач: от обработки фазового пространства сердечного ритма [2] и нормализации параметров нейронных сетей [3] до встраивания цифровых водяных знаков [4-5] и обработки пространств полудифференцируемых функций [6].

Для квантования обычно используется шкала, включающая набор интервалов квантования, внутри каждого из которых выбран представительный уровень. При квантовании все значения, находящиеся внутри интервала, заменяются на уровень квантования, принадлежащий этому интервалу.

Таким образом, результат квантования полностью определяется шкалой. Чаще всего используются равномерные шкалы, в которых интервалы имеют одинаковый размер, а уровни располагаются по центрам интервалов.

Однако, во многих ситуациях, в частности при малом количестве уровней, использование равномерных шкал приводит к недопустимо большой погрешности. В таких ситуациях используются неравномерные шкалы, наиболее известной из которых является шкала Макса. Такие шкалы строятся исходя из минимизации квадратичной погрешности. Ограничением при решении этой задачи оптимизации выступает заданное (фиксированное) количестве уровней квантования.

Несмотря на оптимальность по погрешности, шкалы Макса не являются наилучшими при решении многих прикладных задач, в которых нужно оптимизировать не погрешность, а какой-либо другой показатель качества. Кроме того, количество уровней шкалы при этом может быть неизвестно, что влечёт за собой необходимость использования другого ограничения при оптимизации шкалы, отличного от ограничения на количество уровней.

Например, в задаче компрессии [7-10] часто необходимо минимизировать объём сжатых данных при заданной погрешности. Такие требования приводят к необходимости замены как показателя качества, так и ограничения при решении задачи оптимизации шкалы квантования.

В данной работе проводится обобщение квантователя [10], предложенного в рамках решения задачи компрессии, на случай произвольных критерия и ограничения, используемых при оптимизации шкалы, а также формулируются требования к критерию и ограничению, выполнение которых необходимо для работоспособности алгоритма.

2. Оптимизация квантователя

2.1 Квантователь с неравномерной шкалой

Опишем квантователь с неравномерной [1, 8] шкалой. Будем квантовать целочисленную (для простоты изложения) величину $x \in (L, R]$. Представим $(L, R]$ как объединение интервалов квантования $(b_j, b_{j+1}]$:

$$(L,R] = \bigcup_{j=0}^{N-2} (b_j, b_{j+1}],\ b_j \in Z, b_0 = L-1,\ b_{N-1} = R \quad , \tag{1}$$

где N – количество границ b_j интервалов $(b_j, b_{j+1}]$. Вектор границ интервалов квантования далее обозначим $\mathbf{b}$:

$$\mathbf{b} = \left\{ b_j \in \mathrm{Z} : b_j < b_{j+1}, b_0 = L-1, b_{N-1} = R,\ j \in [0, N) \right\} \tag{2}$$

Внутри интервалов $(b_j, b_{j+1}]$ укажем уровни квантования $c_j \in (b_j, b_{j+1}]$. Вектор уровней квантования далее обозначим $\mathbf{c}$:

$$\mathbf{c} = \left\{ c_j \in \mathrm{Z} : b_j \le c_j \le b_{j+1},\ j \in [0, N-2] \right\} \tag{3}$$

Результатом квантования $\Phi_q(x)$ значения x является уровень квантования c_j, соответствующий интервалу $(b_j, b_{j+1}]$, содержащему x :

$$\Phi_q(x) = c_j : x \in (b_j, b_{j+1}] \tag{4}$$

Требование 1. Пусть существует функция $C(b_j, b_{j+1})$, вычисляющая уровень квантования c_j по соответствующему интервалу $(b_j, b_{j+1}]$

$$c_j = C(b_j, b_{j+1}) \ . \tag{5}$$

В этой ситуации для того, чтобы задать шкалу квантования, достаточно указать вектор $\mathbf{b}$ границ интервалов квантования (и количество N компонент этого вектора). Поэтому далее будем употреблять термин «шкала квантования $\mathbf{b}$ » и «шкала квантования $(b_0, ..., b_{N-1})$ ».

2.2 Постановка задачи оптимизации квантователя

Обозначим $Q(\mathbf{b})$ показатель качества, оптимизируемый (для определённости, минимизируемый) при построении шкалы квантования. Будем проводить оптимизацию с использованием показателя ограничения $E(\mathbf{b})$, который не должен превышать предельное значение $E_{\max}$. Для указания шкалы нужно подобрать количество интервалов N и собственно значения границ интервалов $\mathbf{b}$. Таким образом, задача оптимизации шкалы квантования может быть записана в виде:

$$\begin{cases} Q(\mathbf{b}) \to \min\limits_{N,\mathbf{b}} \\ E(\mathbf{b}) \le E_{\max} \end{cases} \quad (6)$$

Требование 2. Должен существовать способ вычисления показателя качества шкалы $(b_0,...,b_{N-1})$ через показатель качества подшкал $(b_0,...,b_{N-2})$ и (b_{N-2},b_{N-1}). Для простоты изложения (без ограничения общности) далее будем предполагать, что показатели качества таких шкал суммируются (алгоритм останется работоспособным и для более сложных способов вычисления показателей качества):

$$Q(b_0,...,b_{N-1}) = Q(b_0,...,b_{N-2}) + Q(b_{N-2},b_{N-1}). \quad (7)$$

Требование 3. Пусть аналогичное требование выполняется также для показателя ограничения (возможность вычислить показатель ограничения для шкалы через соответствующие показатели подшкал):

$$E(b_0,...,b_{N-1}) = E(b_0,...,b_{N-2}) + E(b_{N-2},b_{N-1}). \quad (8)$$

Нетрудно видеть, что сформулированные требования являются довольно слабыми, так как выполняются в большинстве практических ситуаций.

2.3 Алгоритм оптимизации квантователя

Зададим достаточно малый шаг Δ_E по значению показателя ограничения (параметр алгоритма). Диапазон $[0..E_{\max}]$ значений показателя ограничения разобьём на K поддиапазонов одинакового размера $\Delta_E = E_{\max}/K$.

Шаг №1. Построение оптимальных шкал из двух интервалов.

На всех интервалах $[L,r],\ r\in[L+1,R]$ будем рассчитывать состоящие из двух интервалов квантования оптимальные шкалы, обладающие показателем ограничения $E(\mathbf{b}) = k\Delta_E, k\in[0,K-1]$. Интервалы квантования этих шкал запишем в виде:

$$(b_0,b_1] = (L,d],\ (b_1,b_2] = (d,R],\ L\le d<r, \quad (9)$$

где d – единственнная граница интервала, которую нужно подобрать для каждой искомой шкалы.

Показатель качества шкал из двух интервалов можно записать через значения показателя качества на интервалах:

$$q^{(1)}(r,d) = Q(L,d) + Q(d,r). \quad (10)$$

Аналогичным образом можно записать показатель ограничения шкал из двух интервалов:

$$e^{(1)}(r,d) = E(L,d) + E(d,r). \quad (11)$$

Искомое значение границы $d\in[L,r)$ для шкал из двух интервалов, оптимальных на всех диапазонах $[L,r],\ r\in[L+1,R]$, для каждого значения показателя ограничения $E(\mathbf{b}) = k\Delta_E, k\in[0,K-1]$ может быть найдено перебором и помещено в матрицу $B^{(1)}$:

$$B^{(1)}_{r,k} = \arg\min_{L\le d<r}\left\{q^{(1)}(r,d) : e^{(1)}(r,d)\le k\Delta_E\right\}, r\in[L+1,R], k\in[0,K-1]. \quad (12)$$

Соответствующие значения показателя качества поместим в матрицу $Q^{(1)}$:

$$Q^{(1)}_{r,k} = q^{(1)}(r,B^{(1)}_{r,k}), r\in[L+1,R], k\in[0,K-1].$$

Шаг № j. Построение оптимальных шкал из $(j+1)$ интервала.

На всех интервалах $[L,r],\ r\in[L+1,R]$ будем рассчитывать состоящие из $(j+1)$ интервала квантования оптимальные шкалы, обладающие показателем ограничения $E(\mathbf{b}) = k\Delta_E, k\in[0,K-1]$. Предпоследнюю границу $d\in[L+j-1,r)$ каждой шкалы будем искать перебором.

Для каждого значения границы $d \in [L+j-1, r)$ на предыдущем шаге алгоритма была найдена шкала из j интервалов, оптимальная на интервале $[L,d]$ при значении показателя ограничения $E(\mathbf{b}) = k\Delta_E - E(d,r)$. Таким образом, показатель качества шкал из $(j+1)$ интервала на интервалах $[L,r],\ r \in [L+1, R]$, содержащих оптимальную подшкалу из j интервалов и имеющих показатель ограничения $E(\mathbf{b}) = k\Delta_E, k \in [0, K-1]$, может быть записан в виде:

$$q^{(j)}(r,k,d) = Q^{(j-1)}_{d,t} + q(d,r),\ t = \left[\left(k\Delta_E - E(d,r) \right) / \Delta_E \right]. \qquad (13)$$

Оптимизация этой функции по d позволяет вычислить предпоследнюю границу искомой оптимальной шкалы, которая помещается в матрицу $B^{(j)}$:

$$B^{(j)}_{r,k} = \arg\min_{L \le d < r} q^{(j)}(r,k,d),\ r \in [L+1, R], k \in [0, K-1]. \qquad (14)$$

Соответствующие значения показателя качества поместим в матрицу $Q^{(j)}$:

$$Q^{(j)}_{r,k} = q^{(j)}(r, B^{(j)}_{r,k}), r \in [L+1, R], k \in [0, K-1]. \qquad (15)$$

Алгоритм останавливается на шаге номер $R-L-1$. После этого запускается рекурсивная процедура расчёта искомой оптимальной шкалы.

Рекурсивный расчёт оптимальной шкалы.

Матрицы $Q^{(j)}_{R,K-1}, 0 < j < R-L$ содержат показатель качества шкал из $(j+1)$ интервала, оптимальных на диапазоне $(L,R]$ при значении показателя ограничения $E(\mathbf{b}) = (K-1)\Delta_E$. Оптимизация по j даст номер шага, на котором была построена искомая оптимальная шкала, а количество уровней в ней будет на два больше этого номера шага:

$$N = 2 + \arg\min_{0 < j < R-L} Q^{(j)}_{R,K-1}. \qquad (16)$$

Первая b_0 и последняя b_{N-1} границы этой оптимальной шкалы известны, предпоследняя граница b_{N-2} берется из матрицы B:

$$b_0 = L,\ b_{N-1} = R,\ b_{N-2} = B^{(N-2)}_{R,K-1}, \qquad (17)$$

Остальные границы вычисляются посредством рекуррентной процедуры:

$$b_j = B^{(j)}_{v,w},\ v = b_{j+1},\ w = \left[\left(E_{\max} - \sum_{i=j+1}^{N-2} E(b_i, b_{i+1}) \right) \Big/ \Delta_E \right],\ j = N-3, N-2, \ldots, 1. \qquad (18)$$

На этом построение шкалы завершено. Построенная шкала квазиоптимальна (асимптотически оптимальна при Δ_E, стремящемся к нулю). Показатель ограничения построенной квазиоптимальной шкалы находится в диапазоне $\left[E_{\max} - N\Delta_E, E_{\max} \right]$, так как на каждом шаге отклонение показателя ограничения построенных шкал от оптимальных увеличивается не более чем на Δ_E.

3. Экспериментальное исследование алгоритма оптимизации квантователя

Для исследования эффективности предложенного алгоритма оптимизации квантователя были проведены вычислительные эксперименты. Исследовалась задача построения шкал квантования, оптимальных для задачи компрессии с управляемой погрешностью [7-12], то есть минимизирующих объём сжатых данных при заданной погрешности.

В качестве показателя качества использовалась энтропия $H(\mathbf{b})$ [7] квантованных значений, хорошо приближающая объём сжатых данных, а в качестве показателя ограничения – дисперсия $\varepsilon^2_{MSE}(\mathbf{b})$ ошибки квантования [12]. В результате задача оптимизации (6) решалась в следующей постановке:

$$\begin{cases} Q(\mathbf{b}) = H(\mathbf{b}) = -\sum_{j=0}^{N-2} P(b_j, b_{j+1}) \log_2 P(b_j, b_{j+1}) \to \min_{N,\mathbf{b}} \\ E(\mathbf{b}) = \varepsilon_{MSE}^2(\mathbf{b}) = \sum_{j=0}^{N-2} \sum_{x=b_j}^{b_{j+1}} f(x)\left(x - C\left(b_j, b_{j+1}\right)\right)^2 \le \varepsilon_{\max}^2 \end{cases} \tag{19}$$

где $f(x)$ это плотность распределения вероятностей квантуемой величины x (погрешностей предсказания отсчётов сжимаемого сигнала), а вероятность $P(l,r)$ попадания x в интервал $\left(l,r\right]$ и функция вычисления уровня квантования $C(l,r)$ заданы следующим образом:

$$P(l,r) = \sum_{x=l+1}^{r} f(x), \qquad C(l,r) = \sum_{x=l+1}^{r} x\, f(x) \Big/ \sum_{x=l+1}^{r} f(x), \tag{20}$$

где уровень $C(l,r)$ фактически вычисляется как локальное среднее по интервалу квантования.

В качестве плотности распределения использовалась функция вида $f(x) = \exp\left(-\xi|x|\right)/\left(2\xi\right)$, характерная для квантуемых сигналов (погрешностей предсказания отсчётов) дифференциальных [11], иерархических [9] и многих других [7, 12] методов компрессии.

В качестве базы для сравнения использовались равномерные шкалы и шкалы Макса [7]. Шкалы Макса строились исходя из минимизации квадратичной погрешности квантования $\varepsilon_\Sigma^2(\mathbf{b})$ при ограничении на количество уровней квантования $N = N_{\max}$:

$$\begin{cases} \varepsilon_\Sigma^2(\mathbf{b}) = \sum_{j=0}^{N-2} \sum_{x=b_j}^{b_{j+1}} f(x)\left(x - C\left(b_j, b_{j+1}\right)\right)^2 \to \min_{\mathbf{b}} \\ N = N_{\max} \end{cases} \tag{21}$$

Типичные результаты показаны на рис. 1, на котором показаны графики зависимости энтропии квантованных ошибок интерполяции от квадратичной погрешности интерполяции. Нетрудно видеть, что предложенный алгоритм позволяет строить шкалы, имеющее преимущество в координатах «погрешность-энтропия» над равномерными шкалами и шкалами Макса (обеспечивает меньший объем сжатых данных при той же погрешности и, соответственно, меньшую погрешность при совпадающем объёме сжатых данных), причём это преимущество увеличивается с ростом погрешности.

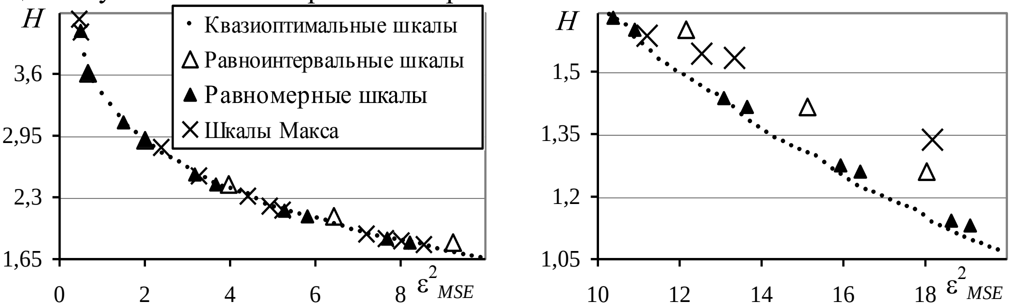

Рисунок 1. Исследование алгоритма построения оптимальных шкал квантования.

4. Заключение

Рассмотрена задача построения шкал квантования, оптимальных по заданному критерию и удовлетворяющих выбранному ограничению. Рассмотрена формальная постановка такой задачи оптимизации. Предложен алгоритм построения квазиоптимальных шкал квантования, аппроксимирующих оптимальные шкалы с заданной точностью при соблюдении ограничения. Сформулированы требования к критерию оптимизации и ограничению, обеспечивающие работоспособность алгоритма. Проведены вычислительные эксперименты по построению квазиоптимальных шкал. Полученные результаты экспериментов подтвердили преимущество построенных шкал квантования над известными.

5. Благодарности
Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 18-01-00667, а также Министерства науки и высшего образования РФ в рамках выполнения работ по Государственному заданию ФНИЦ «Кристаллография и фотоника» РАН (соглашение № 007-ГЗ/Ч3363/26).

6. Литература

[1] Оппенгейм, А. Цифровая обработка сигналов / А. Оппенгейм, Р. Шафер – Litres, 2018. – 1048 с.
[2] Кудинов, А.Н. Квантование фазового пространства мгновенного сердечного ритма / А.Н. Кудинов, С.А. Михеев, В.П. Цветков, И.В. Цветков // Доклады Международной конференции "Математическая биология и биоинформатика". – 2019. – Т. 7. – С. e15. DOI: 10.17537/icmbb18.21.
[3] Колесников, Д.С. Оптимальное равномерное квантование параметров сверточных нейронных сетей / Д.С. Колесников, Д.А. Кузнецов // Вопросы радиоэлектроники. – 2018. – Т. 8. – С 99-103. DOI: 10.21778/2218-5453-2018-8-99-103.
[4] Евсютин, О.О. Стеганографическое встраивание дополнительных данных в снимки дистанционного зондирования Земли с помощью метода QIM с переменным шагом квантования в частотной области / О.О. Евсютин, А.С. Кокурина, Р.В. Мещеряков // Известия Томского политехнического университета инженерных георесурсов. – 2019. – Т. 330, № 8. – С. 155-162.
[5] Егорова, А.А. Классификация и сравнительное исследование систем аутентификации JPEG-изображений, основанных на встраивании полухрупких водяных знаков / А.А. Егорова, В.А. Федосеев // Компьютерная оптика. – 2019. – Т. 43, № 3. – С. 419-433. DOI: 10.18287/2412-6179-2019-43-3-419-433.
[6] Сергеев, А.Г. Квантование Соболевского пространства полудифференцируемых функций // Матем. сб. – 2016. – Т. 207, № 10. – С. 96-104. DOI: https://doi.org/10.4213/sm8673.
[7] Pearlman, W.A. Digital Signal Compression: Principles and Practice / W.A. Pearlman, A. Said — Cambridge University Press, 2011. – 419 p.
[8] Сойфер, В.А. Методы компьютерной обработки изображений – М.: Физматлит, 2003. – 784 с.
[9] Sergeyev, V.V. Compression Method for Real-Time Systems of Remote Sensing / V.V. Sergeyev, N.I. Glumov, M.V. Gashnikov // Proceedings of 15th International Conference on Pattern Recognition. – 2000. – Vol. 3. – P. 232-235.
[10] Гашников, М.В. Оптимальное квантование в задаче компрессии цифровых сигналов // Компьютерная оптика. – 2001. – Т. 21. – С. 179-184.
[11] Максимов, А.И. Адаптивная интерполяция многомерных сигналов при дифференциальной компрессии / М.В. Гашников, А.И. Максимов // Компьютерная оптика. – 2018. – Т. 42, № 4. – С. 679-687. DOI: 10.18287/2412-6179-2018-42-4-679-687.
[12] Sayood, K. Introduction to Data Compression – The Morgan Kaufmann Series in Multimedia Information and Systems, 2012. – 743 p.

Algorithm for optimizing quantization scales by an arbitrary quality measure

N.I. Glumov[1,2], M.V. Gashnikov[1,2]

[1]Samara National Research University, Moskovskoe Shosse 34A, Samara, Russia, 443086
[2]Image Processing Systems Institute of RAS - Branch of the FSRC "Crystallography and Photonics" RAS, Molodogvardejskaya street 151, Samara, Russia, 443001

Abstract. The paper considers the problem of constructing quantization scales that are optimal according to a specified criterion and satisfy a specified constraint. The mathematical statement of the optimization problem is considered. An algorithm is proposed for constructing quasi-optimal quantization scales that approximate the optimal scales with specified accuracy, subject to the constraint. Requirements to the optimization criterion and the constraint are formulated, which ensure the operability of the algorithm. Computational experiments are being carried out to construct quasi-optimal scales. The experimental results confirm the advantage of the constructed scales over the known ones.