Анализ распознавания одномерных сигналов в шуме.

Петюшко А.А., Мазуренко И.Л.
Каф. Математической теории интеллектуальных систем
механико-математического факультета МГУ им. М.В. Ломоносова
$e-mail: petsan@newmail.ru, ivanmaz@mech.math.msu.su$

1. Введение
Необходимые сведения и основные определения

Любое колебание можно охарактеризовать частотой и фазой, точнее, спектром как зависимостью амплитуды и фазы колебания от частоты (для данного момента времени). Полезному сигналу обычно сопутствуют иные источники сигнала, которые можно в совокупности назвать источниками шума. Шум накладывается на полезный сигнал по аддитивному закону (в амплитудно-временной области).

Одномерный сигнал $s(t) : \mathbb{R}_+ \to \mathbb{R}$ можно рассматривать как непрерывную функцию зависимости амплитуды (как величины со знаком) от времени. Однако, при цифровой обработке сигнал подвергается дискретизации в соответствии с выбранной частотой, т.е. производится запись значений амплитуды через определенные промежутки времени.

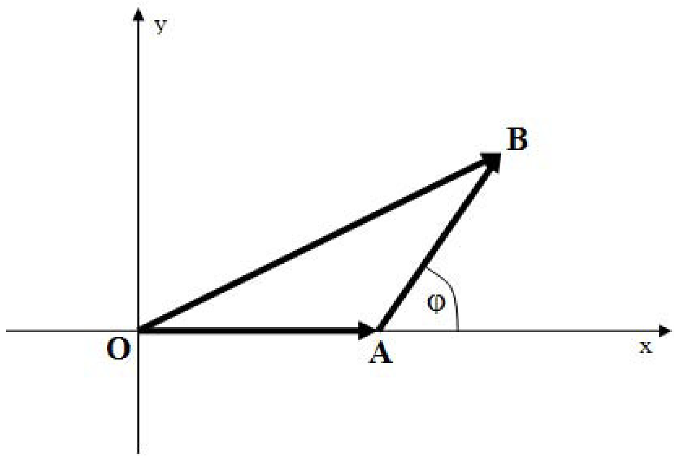
Чтобы по дискретной последовательности значений амплитуды восстановить сигнал, используется так называемое ***дискретное преобразование Фурье (ДПФ)***, которое текущему окну сигнала длины N ставит в соответствие вектор коэффициентов разложения в ряд Фурье $c_0, ..., c_{N-1}$, причем $c_j = \overline{c_{N-j}}, \quad j = \overline{N/2+1, N-1}$, т.к. значения амплитуд действительны.

Назовем *"вектором Фурье"* элемент множества $\mathbb{R}_+^{N/2+1}$, составленный из модулей первых $N/2+1$ коэффициентов в разложении в ряд Фурье. Порядковый номер коэффициента в векторе Фурье определяет частоту сигнала.

Геометрическая модель зашумления

Фактически каждый элемент из полученного вышеописанным образом вектора Фурье представляет собой коэффициент (амплитуду) при синусоиде, соответствующей некоторой частоте. Таким образом, после ДПФ двух сигналов и взятия модуля от коэффициентов мы имеем амплитуды синусоид сигналов на соответствующих частотах, причем фазы у синусоид каждого преобразования, вообще говоря, свои.

Значит, чтобы сложить два сигнала, имея их вектора Фурье, нам фактически нужно сложить (по каждой частоте отдельно) две синусоиды с разными фазами и амплитудами, что геометрически эквивалентно сложению двух векторов с длинами, равными амплитудам синусоид, и углом между ними, равным разности фаз:



В данном случае на рисунке изображено сложение двух векторов $\overrightarrow{OA}$ и $\overrightarrow{AB}$, соответствующих, например, амплитудам речи и шума, с разностью фаз между соответствующими векторами φ. Таким образом, в данной модели за зашумленный сигнал (например, зашумленная речь) мы принимаем $\overrightarrow{OB}$.

Назовем это *"геометрической моделью зашумления"*.

Распознавание одномерных сигналов. Методы распознавания в условиях шума

Часто, прежде всего для практических нужд, встает задача ***распознавания звуковых сигналов*** [1] - т.е. определения, какому типу (классу) принадлежит исследуемый звуковой сигнал. Деление на классы может быть совершенно различным, будь то классы полезный сигнал/шум, голос определенного диктора или конкретный сигнал из какого-либо словаря.

Одним из основных методов распознавания является ***метод сравнения с эталонами*** (эталон - заранее записанный образец того или иного полезного сигнала). Он основан на том, что на основе выбранной метрики (расстояния) между одномерными сигналами результатом распознавания сигнала считается значение того эталона, который *ближе* (расстояние от сигнала до которого *меньше*) к рассматриваемому сигналу.

Традиционно применяются следующие методы распознавания в шуме, основанные на методе сравнения с эталонами:

1. Использование эталонов, записанных в условиях отсутствия шума. Самый простой метод.

2. ***"Метод спектрального вычитания"*** [2].

Сначала вычитаем из амплитуды на каждой частоте оценку амплитуды шума, затем сравниваем полученный сигнал с эталоном.

3. ***"Метод адаптивной фильтрации"*** (обучения в шуме) [3], [4].

На каждом шаге производим настройку параметров адаптивного фильтра (например, с помощью метода наименьших квадратов), используя признак "сигнал/шум".

4. ***"Метод маскирования шума"*** [5].

На каждой отдельной частоте путем применения введенной выше "геометрической модели зашумления" зашумляем эталон с шумом со случайной фазой. Затем сравниваем распознаваемый сигнал с "зашумленным" эталоном.

Вышеперечисленные методы распознавания используются не только в распознавании речи (что было взято за основу для работы), но и в ядерно-магнитном резонансе (ЯМР), химической акустике и других отраслях науки [6], где используется очистка одномерных сигналов от шума.

2. Сравнение методов распознавания сигнала (сравнения с эталоном) в условиях шума

Сделанные допущения

Без ограничения общности будем рассматривать сигнал на какой-нибудь фиксированной частоте. Все выкладки очевидным образом распространяются на все частоты.

Полагаем, что разность фаз для пары сигнал-шум на этой частоте - это случайная величина φ, равномерно распределенная на отрезке $[0, 2\pi]$.

Без ограничения общности в качестве эталонов будем рассматривать 2 неотрицательных числа $0 \leq e_1 < e_2$.

Для простоты положим, что амплитуды незашумленного сигнала и шума - две случайные равномерно распределенные величины на отрезках $[a, b]$ и $[c, d]$ соответственно: $s \in [a, b]$ - сигнал и $n \in [c, d]$ - шум, значения которых принадлежат $\mathbb{R}_+$ (т.е. $a \geq 0$ и $c \geq 0$). Для сравнения методов 1) и 2) примем $a = e_1, b = e_2$.

Случайные величины s, n, φ - независимы.

Также полагаем, что ни при каком значении разности фаз сигнал-шум первый эталон не станет дальше от нуля, чем второй (т.е. человек способен будет еще их различить). Это порождает условие на максимальную амплитуду шума, которая не может быть больше половины расстояния между эталонами, т.е. $n \leq \frac{L}{2}$, где $L = e_2 - e_1$.

Для сравнения методов 1) и 2) предположим, что расстояние между эталонами меньше расстояния от первого эталона до нуля, т.е. $L < e_1$.

Постановка задач

В рамках вышеперечисленных допущений необходимо решить следующие задачи:

A. Пусть $s \oplus n$ - зашумленный сигнал (см. "геометрическую модель зашумления" - там фигурировал вектор $\overrightarrow{OB}$). Требуется найти распределение $p(s \oplus n)$.

B. Пусть α - знак неравенства ($<$ или $>$). Далее, пусть имеем изначально неравенство $\rho(s, e_1)\ \alpha\ \rho(s, e_2)$. Требуется найти вероятность правильного распознавания, т.е. вероятность того, что после преобразования (которого, вообще говоря, может и не быть - например, метод 1) сигнала и эталонов $s \to s', e_1 \to e_1', e_2 \to e_2'$ сигнал не станет ближе к другому, т.е.

$$\mathbf{P}_0 = \mathbf{P}(\rho(s, e_1')\ \alpha\ \rho(s, e_2') \bigcup \rho(s, e_1)\ \alpha\ \rho(s, e_2))$$

C. Пусть $\mathbf{P}_0^i$ - вероятность правильного распознавания для метода с номером i. Получить аналитические формулы для $\mathbf{P}_0^1, \mathbf{P}_0^2, \mathbf{P}_0^3$.

D. Для некоторых значений s и n выяснить знак α в неравенстве $\mathbf{P}_0^1\ \alpha\ \mathbf{P}_0^2$.

Вычисление плотности распределения $s \oplus n$

Теорема 1. *Плотность распределения $s \oplus n$ вычисляется по формуле*

$$p(s \oplus n) = \frac{1}{\pi(b-a)(d-c)} \int\limits_a^b \int\limits_c^d \frac{s \oplus n}{sn\sqrt{1 - (\frac{s^2+n^2-s\oplus n^2}{2sn})^2}} dnds.$$

Доказательство. Используются теоремы о замене переменных под знаком тройного интеграла и о плотности совместного распределения независимых случайных величин (s, n, φ). Также считаем, что разность фаз равномерно распределена на отрезке $[0, \pi]$. ■

Вычисление вероятности правильного распознавания методом сравнения с эталоном в условиях шума

Пусть имеется два эталона p и $q, p < q$, и сигнал r. Преобразованные (зашумленные/расшумленные/оставленные без изменений) эталоны и сигнал - p', q' и r' соответственно.

Утверждение. *Вероятность правильного распознавания вычисляется по формуле*

$$\mathbf{P}_0 = \mathbf{P}(r < \frac{p+q}{2}, r' < \frac{p'+q'}{2}) + \mathbf{P}(r \geq \frac{p+q}{2}, r' \geq \frac{p'+q'}{2}). \qquad (0)$$

Доказательство. Очевидно, что близость сигнала r к соответствующему эталону будет определяться его положением относительно середины отрезка $[p, q]$, аналогично и для p', q' и r'. Вероятность правильного распознавания - это вероятность того, что после преобразования сигнал не стал ближе к другому эталону. В итоге - искомая формула. ■

Вывод аналитических формул для расчета вероятности правильного распознавания для методов 1), 2) и 3)

Для метода 1) имеем $p = p' = e_1, q = q' = e_2, r = s, r' = s \oplus n$. Значит, по формуле (0) получим:

$$\mathbf{P}_0^1 = \frac{1}{\pi(b-a)(d-c)} \times \left(\int\limits_{|s-n|}^{\frac{e_1+e_2}{2}} \int\limits_{a}^{\frac{e_1+e_2}{2}} \int\limits_{c}^{d} \frac{s \oplus n}{sn\sqrt{1-(\frac{s^2+n^2-s\oplus n^2}{2sn})^2}} dndsd(s\oplus n) + \right.$$

$$\left. + \int\limits_{\frac{e_1+e_2}{2}}^{s+n} \int\limits_{\frac{e_1+e_2}{2}}^{b} \int\limits_{c}^{d} \frac{s \oplus n}{sn\sqrt{1-(\frac{s^2+n^2-s\oplus n^2}{2sn})^2}} dndsd(s\oplus n) \right). \qquad (1)$$

Следовательно, верна

Лемма. *Вероятность правильного распознавания методом 1) выражается формулой (1).*

Вывод формул для случаев 2) и 3) совершенно аналогичен (меняются только пределы интегрирования и увеличивается количество слагаемых).

Сравнение методов 1) и 2)

Воспользуемся сделанными выше допущениями, что $a = e_1, b = e_2$ и $n \leq \frac{L}{2}$, где $L = e_2 - e_1$. Теперь обозначим за a - середину отрезка $[e_1, e_2]$, т.е. $a = \frac{e_1+e_2}{2}$. За x обозначим расстояние между сигналом s и серединой отрезка $[e_1, e_2]$, т.е. $x = |s-a|$.

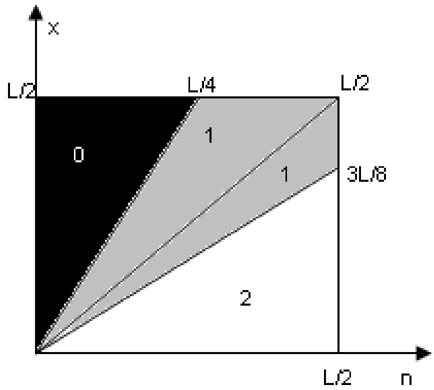
Теорема 2. *При $x \geq 2n$ оба методы не работают.*
При $\frac{3}{4} \leq x < 2n$ лучше работает метод 1).
Доказательство. Преобразуем выражение

$$\mathbf{P}_0^1 - \mathbf{P}_0^2 = \mathbf{P}(s \geq a, s\oplus n \in [a, n+a]) - \mathbf{P}(s < a, s \oplus n \in [a, n+a]). \qquad (2)$$

Для оценки формулы (2) поступаем следующим образом: берем всевозможные $x \in [0, L/2]$, для каждого конкретного x рассматриваем сигналы $s_1 = a - x$ и $s_2 = a + x$, берем конкретное значение $n \in [0, L/2]$, для них вычисляем величину $s \oplus n_1$ и $s \oplus n_2$, и и сравниваем меру λ_1 (под мерой понимаем лебегову меру отрезка - длину) множества углов, задающих разность фаз, удовлетворяющих условию $s \oplus n_2 \in [\frac{e_1+e_2}{2}, n + \frac{e_1+e_2}{2}]$, с мерой λ_2 множества углов, удовлетворяющих $s \oplus n_1 \in [\frac{e_1+e_2}{2}, n + \frac{e_1+e_2}{2}]$.

В том случае, когда $\lambda_1 \geq \lambda_2$, считаем, что выполняется условие $У$. Если $\lambda_1 < \lambda_2$, то считаем, условие $У$ не выполняется. Применяя несложные геометрические построения и необходимые выкладки, находим, что условие $У$ выполняется при $\frac{3}{4} \leq x < 2n$, а при $x \geq 2n$ $\lambda_1 = \lambda_2 = 0$, т.е. условие теоремы выполнено. ■

Замечание. *Мера на множестве допустимых значений (s, n), где метод 1) работает лучше, больше соответствующей меры для метода 2).*
Доказательство.



На рисунке: 0 - где мера допустимых параметров равна нулю, т.е. оба метода не работают, 1 - где выполняется условие $У$, т.е. метод 1) работает лучше, и 2 - где лучше работает второй метод. Мера, где условие $У$ выполняется, есть $\mu_1 > \frac{1}{4}\frac{3L^2}{8}$. Условие $У$ не выполняется на мере $\mu_2 < \frac{1}{4}\frac{3L^2}{8}$. Значит, $\mu_1 > \mu_2$. ∎

3. Благодарности

Авторы благодарят академика Кудрявцева В.Б. и проф. Бабина Д.Н. за внимание к работе и ценные указания.

4. Литература

[1] Ю.И. Журавлев, "Об алгебраическом подходе к решению задач распознавания и классификации". "Проблемы кибернетики", вып. 33, Москва, "Наука", 1978.

[2] S.F. Boll, "Suppression of acoustic noise in speech using spectral subtraction". IEEE, Transactions on Acoustics, Speech, and Signal Processing, vol. ASSP - 27, no. 2, pp. 113-120, 1979.

[3] B. Widrow et al., "Adaptive noise cancelling: Principles and applications". Proceedings of the IEEE, 63(12), pp. 1692-1716, 1975.

[4] L. Bamber, D. Kerns, S. Haykin, "Adaptive Filter Theory" (International edition). Prentice Hall, 2004.

[5] D.H. Klatt, "A digital filter bank for spectral matching". Proc. ICASSP, 76, pp.573-576, 1976.

[6] A.E. Derome, "Modern NMR techniques for chem. research Pergamon Press, November 1993.