Техника исследования гравитационных волн

Лекция 10

Информационно-энтропийный подход к фильтрации сигналов

Усіпов Н.М., PhD · КазНУ имени аль-Фараби

Цель и план лекции

Цель

Изучение способов использования энтропийных критериев для фильтрации зашумлённых сигналов, принципов выбора порога и базиса, а также методов оценки качества и информативности результата фильтрации.

1. Введение
2. Принципы энтропийной фильтрации
3. Выбор порога на основе энтропии
4. Выбор базиса и параметров
5. Оценка результата
6. Классические методы пороговой фильтрации
7. Энтропийные функционалы и выбор представления
8. Итеративная фильтрация и восстановление сигнала

1 Введение

- При фильтрации всегда возникает вопрос: какие компоненты следует оставить, а какие удалить?
- Такая стратегия особенно удобна при неточно известной форме сигнала. Однако следует понимать ограничение: критерий определяет структуру, но не её физическое происхождение, поэтому глитч также может быть сохранён как компактная компонента.

2 Принципы энтропийной фильтрации

Разреженное представление

Сигнал преобразуется в базис, где его энергия концентрируется в малом количестве коэффициентов, например с помощью кратковременного преобразования Фурье или вейвлет-преобразования.

Оценка распределения

Нормируются значения $|c_k|^2$, образуя распределение энергии по коэффициентам.

Энтропийный критерий

Определяется порог, число компонент или параметры преобразования.

Реконструкция

Отобранные коэффициенты преобразуются обратно во временную область.

2 Принципы энтропийной фильтрации (продолжение)

Контроль качества

Оцениваются показатели сохранения сигнала и удаления шума.

3 Выбор порога на основе энтропии

Значение порога T определяет компромисс: слишком низкий порог оставляет шум, а слишком высокий – разрушает сигнал.

Метод Капура и др. (1985)

Значения разделяются на два класса (шум и сигнал) таким порогом, при котором сумма энтропий двух классов максимальна:

$$T^* = arg\ max\ [\ H_0(T) + H_1(T)\ ].$$

Обобщённые энтропии

В выражение подставляются энтропии Реньи или Тсаллиса; порядок становится настраиваемым параметром.

Сравнение со статистикой шума

Коэффициенты, распределение которых не отличается от распределения шума, отбрасываются, что выражается с помощью дивергенции Кульбака–Лейблера.

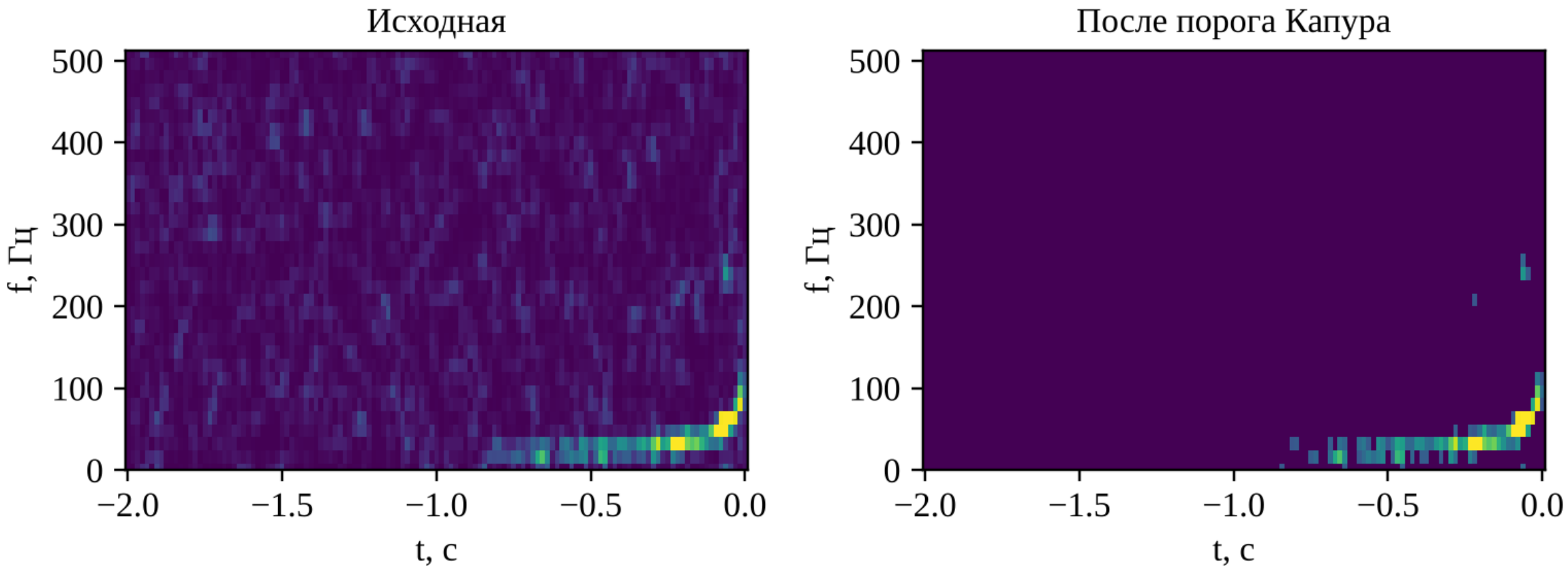
Минимальная длина описания

Выбирается модель, кратко кодирующая данные и позволяющая отделить шум.

4 Выбор базиса и параметров

- Энтропия также служит критерием выбора представления. Алгоритм наилучшего базиса Койфмана и Викерхаузера (1992) среди вейвлет-пакетов выбирает базис с минимальной суммарной энтропией.
- Необходимо также учитывать, что минимизация энтропии сама по себе не является целью: если обнулить все коэффициенты, кроме одного, энтропия станет нулевой, но сигнал будет разрушен.

4 Выбор базиса и параметров (продолжение)



Спектрограмма сигнала до и после порога Капура

5 Оценка результата

- Для синтетических данных с известным сигналом применяются следующие показатели:
- коэффициент корреляции с исходным сигналом;
- среднеквадратичная ошибка и отношение сигнал/шум после фильтрации;
- относительная доля оставшейся энергии сигнала;
- фазовая ошибка восстановленного сигнала.
- Для реальных данных истинный сигнал неизвестен, поэтому качество оценивают косвенно: инъекцией модельного сигнала в реальный шум и по статистической проверке остатка, который должен соответствовать свойствам шума.

5 Фильтрация с порогом Капура

Листинг 10.1.

```
import numpy as np
from scipy import signal

def chirp(Mc, fs, T, flow=20):
    m = Mc * 4.925e-6
    t = np.arange(-T, 0, 1 / fs)
    tau = 1 / fs - t
    f = (5 / (256 * tau))**0.375 / (np.pi * m**0.625)
    ph = -2 * (5 * m)**-0.625 * tau**0.625
    h = (np.pi * m * f)**(2 / 3) * np.cos(ph)
    h[(f < flow) | (f > 3830 / Mc)] = 0
    return t, h / np.sqrt((h**2).sum())
```

5 Фильтрация с порогом Капура (продолжение)

Листинг 10.1.

```
def kapur(P, bins=256):
    hist, edges = np.histogram(P, bins)
    q = hist / hist.sum()
    c = q.cumsum()
    best, kbest = -np.inf, 0
    for k in range(1, bins - 1):
        if c[k] <= 0 or c[k] >= 1:
            continue
        p0 = q[:k + 1][q[:k + 1] > 0] / c[k]
        p1 = q[k + 1:][q[k + 1:] > 0] / (1 - c[k])
        H = -(p0 * np.log(p0)).sum() - (p1 * np.log(p1)).sum()
        if H > best:
            best, kbest = H, k
    return edges[kbest + 1]

def corr(a, b):
    return (a @ b) / np.sqrt((a @ a) * (b @ b))

fs, T = 4096, 2.0
t, h = chirp(28, fs, T)
kw = dict(fs=fs, nperseg=256, noverlap=192)
```

5 Фильтрация с порогом Капура (продолжение)

Листинг 10.1.

```
print("SNR  r(исх.)  r(фильтр)  доля коэфф.")
for snr in (15, 25, 40):
    rng = np.random.default_rng(snr)
    x = rng.normal(size=t.size) + snr * h
    f, tt, Z = signal.stft(x, **kw)
    P = np.abs(Z)**2
    Tk = kapur(P)
    _, y = signal.istft(Z * (P > Tk), **kw)
    y = y[:x.size]
    print(f"{snr:3} {corr(x, h):8.3f} {corr(y, h):10.3f}"
          f" {(P > Tk).mean():11.4f}")
```

5 Результат: Фильтрация с порогом Капура

SNR	r(исх.)	r(фильтр)	доля коэфф.
15	0.154	0.448	0.0073
25	0.271	0.861	0.0052
40	0.398	0.962	0.0080

Результат показывает значительное повышение коэффициента корреляции при малой доле сохранённых коэффициентов.

6 Классические методы пороговой фильтрации

- Вейвлетная пороговая фильтрация, разработанная Донохо и Джонстоном, основана на разделении коэффициентов на значимые и шумовые.

$$T = \sigma \sqrt{(2 \ln N)}.$$

- Жёсткий порог обнуляет коэффициенты с модулем меньше T и оставляет остальные без изменения.
- Минимаксный порог и критерий SURE (несмещённая оценка риска Стейна) выбирают значение на основании минимизации ожидаемой ошибки.

7 Энтропийные функционалы и выбор представления

- Для оценки концентрации распределения применяются различные функционалы. Энтропия Шеннона чувствительна к распределению малых значений; норма l^p (сумма $|c|^p$) при малых p акцентирует разреженность; энтропия Реньи третьего порядка применяется к частотно-временным распределениям.
- Наилучший базис выбирается снизу вверх: для каждого узла дерева вейвлет-пакетов сравнивается аддитивная стоимость родительского узла и сумма стоимостей его потомков.

8 Итеративная фильтрация и восстановление сигнала

- После грубой первичной оценки сигнала можно уточнять оценку шума и порога. Алгоритм итеративно разделяет компоненты: оценённые сигнальные коэффициенты исключаются при вычислении параметров шума, порог пересчитывается, процедура повторяется до стабилизации.
- Проблема восстановления сигнала из частотно-временного представления связана с тем, что после обнуления коэффициентов матрица в общем случае не является точным STFT какого-либо сигнала.
- Возможен учёт физических ограничений: допускаются только те траектории, у которых частота возрастает со временем, как у сигнала сближения, причём скорость определяется чирп-массой в заданном диапазоне.
- Также необходимо контролировать влияние фильтрации на последующие этапы. Если результат применяется для оценки параметров, то нелинейная обработка меняет статистику шума, и обычная функция правдоподобия больше не применима.

9 Разбор примера

1. Пусть распределение энергии по четырём коэффициентам равно p = (0,7; 0,1; 0,1; 0,1).

2. Энтропия равна H = −0,7 ln 0,7 − 3·0,1 ln 0,1 = 0,250 + 0,691 ≈ 0,940 нат, а нормированная величина составляет 0,940/ln 4 ≈ 0,678.

3. Если оставить только первый коэффициент, энтропия станет равной нулю, но будет потеряно 30 % энергии.

4. Если эти коэффициенты относятся к шуму, операция полезна; если там находится часть сигнала, то она вредна.

5. Отсюда следует необходимость дополнительного критерия.

6. Для N = 4096 отсчётов при σ = 1 универсальный порог составляет √(2 ln 4096) ≈ 4,08.

10 Вопросы для обсуждения на занятии

Выбор порога

Какие последствия имеет завышенный и заниженный порог при малом отношении сигнал/шум?

Параметры

Можно ли оценивать чирп-массу по результатам нелинейно отфильтрованного сигнала?

Глитчи

Как отличить компактную структуру сигнала от компактной помехи?

Контрольные вопросы

1. В чём заключается основная идея энтропийной фильтрации?
2. Опишите метод выбора порога Капура.
3. Как можно использовать энтропию для выбора базиса преобразования?
4. Почему минимальное значение энтропии само по себе не является целью?
5. Как глитчи влияют на применение энтропийных критериев?
6. Какие показатели применяются для оценки качества фильтрации?
7. Как оценить качество фильтрации реальных данных при неизвестном сигнале?
8. В чём заключаются ограничения энтропийной фильтрации при малом отношении сигнал/шум?

Список литературы

1. Kapur J.N., Sahoo P.K., Wong A.K.C. A new method for gray-level picture thresholding using the entropy of the histogram // Computer Vision, Graphics, and Image Processing. – 1985. – Vol. 29, No. 3. – P. 273–285.
2. Coifman R.R., Wickerhauser M.V. Entropy-based algorithms for best basis selection // IEEE Transactions on Information Theory. – 1992. – Vol. 38, No. 2. – P. 713–718.
3. Donoho D.L., Johnstone I.M. Ideal spatial adaptation by wavelet shrinkage // Biometrika. – 1994. – Vol. 81, No. 3. – P. 425–455.
4. Rissanen J. Modeling by shortest data description // Automatica. – 1978. – Vol. 14, No. 5. – P. 465–471.
5. Shannon C.E. A mathematical theory of communication // Bell System Technical Journal. – 1948. – Vol. 27. – P. 379–423, 623–656.
6. Abbott B.P., et al. A guide to LIGO–Virgo detector noise and extraction of transient gravitational-wave signals // Classical and Quantum Gravity. – 2020. – Vol. 37, No. 5. – Art. 055002. – DOI: 10.1088/1361-6382/ab685e.