Секция: *Обработка сигналов и изображений*
УДК: 004.93; 519.688; 517.518

Банк линейных фильтров и скалярные нелинейности как универсальная модель на компактных классах сигналов

Крылов А. С.[1*]

[1*]Факультет ВМК, Московский государственный университет имени М.В. Ломоносова.

Автор(ы), ответственный(ые) за переписку. E-mail(s): kryl@cs.msu.ru;

Аннотация

Мы рассматриваем задачу аппроксимации непрерывных скалярных функционалов на классах сигналов и доказываем теорему универсальной аппроксимации на компактах в гильбертовых пространствах. Любой непрерывный функционал на компактном классе многоканальных сигналов может быть равномерно приближён моделью, которая (i) вычисляет конечное число линейных измерений (выходов банка фильтров/проекций), (ii) пропускает их через скалярные непрерывные нелинейности и (iii) агрегирует результат суммированием.

Доказательство использует компактность класса сигналов: сначала множество аппроксимируется конечномерной ортогональной проекцией, затем применяется классическая универсальная аппроксимация в конечномерном пространстве, после чего полученная конструкция переносится обратно в исходное пространство сигналов.

В инженерных терминах результат даёт теоретическое обоснование архитектуры “filterbank + pointwise nonlinearity + pooling” как универсальной модели на компактных классах сигналов (аудио, сейсмика и др.).

Ключевые слова: универсальная аппроксимация, цифровая обработка сигналов, банк фильтров, функции Эрмита, функции Лагерра, эмпирическая модовая декомпозиция, нейронные операторы

Получено редакцией 16.02.2026; внесены авторские правки 17.02.2026; принята к публикации 17.02.2026

1 Введение

Многие задачи цифровой обработки сигналов (аудио, сейсмика, радиосвязь) естественно формулируются как аппроксимация *скалярных функционалов* на пространствах сигналов: по входу $x = (x_1, \dots, x_n)$, где каждый канал x_i лежит в гильбертовом пространстве H, требуется вычислить значение $f(x) \in \mathbb{R}$. В работе рассматриваются непрерывные отображения $f\colon K \to \mathbb{R}$, заданные на компактном классе сигналов $K \subset H^n$. Такая постановка охватывает детектирование (есть/нет), оценивание параметров (задержки, частоты, уровня шума) и построение непрерывных оценок/логарифмов отношения шансов для классификации.

Замечание (про непрерывность f и пороговые решения). Во многих инженерных задачах итоговое решение действительно является пороговым (жёсткий детектор, индикатор события), то есть как отображение сигнала оно разрывно. На практике, однако, обычно строят *непрерывную* "оценку уверенности"/оценочную функцию (например, выход согласованного фильтра, отношение правдоподобий, логит или вероятность), а порог применяют уже после этого. Теорема применима именно к таким непрерывным аппроксимациям f (soft-детекторам) и, соответственно, обосновывает универсальность для сглаженных/непрерывных критериев до стадии порогования.

С практической точки зрения во многих инженерных конвейерах обработки и современных нейросетевых архитектурах используется общий вычислительный шаблон: (i) извлечь конечное число *линейных измерений* сигнала (выходы банка фильтров, проекции на базис, согласованная фильтрация, коэффициенты STFT/вейвлет-преобразования), (ii) применить к каждому измерению *скалярную нелинейность*, (iii) агрегировать результаты (суммирование/пулинг или небольшой нелинейный микшер). Несмотря на распространённость этого подхода, остаётся принципиальный вопрос: достаточно ли такого класса моделей, чтобы описывать произвольные непрерывные зависимости от сигнала (по крайней мере на реалистичных классах входов)?

В настоящей статье мы даём положительный ответ на уровне теоремы универсальной аппроксимации: любая непрерывная функция f на компактном множестве $K \subset H^n$ может быть равномерно приближена конечной суммой вида

$$\sum_{j=1}^{r} \zeta_j\Big(\sum_{i=1}^{n} \varphi_{ji}(x_i)\Big),$$

где $\varphi_{ji} \in H^*$ — непрерывные линейные функционалы (измерения), а $\zeta_j\colon \mathbb{R} \to \mathbb{R}$ — непрерывные скалярные нелинейности. В терминах цифровой обработки сигналов семейство $\{\varphi_{ji}\}$ интерпретируется как набор каналов анализа (фильтры/сенсоры/проекции), а $\{\zeta_j\}$ — как статические нелинейности и итоговая агрегация.

Ключевая идея доказательства состоит в том, что *компактность* класса сигналов позволяет свести задачу к конечномерному заместителю: сначала множество K равномерно аппроксимируется конечномерной ортогональной проекцией в H^n; затем применяется классическая универсальная аппроксимация в $\mathbb{R}^m$ (см. [1, 2]); после чего полученная конструкция переносится обратно в исходное пространство. Такой взгляд делает результат «читабельным для инженера»: конечное число измерений оказывается достаточным не потому, что исходное пространство малоразмерно, а потому, что рассматриваемый класс входов компактен.

Полученная теорема даёт строгую основу для архитектуры “filterbank + pointwise nonlinearity + pooling” как универсальной модели на компактных классах сигналов; по смыслу она близка к современным подходам обучения операторов, где явно выделяются стадия измерения и стадия нелинейного смешивания.

2 Основной результат

Теорема 1 (аппроксимация на компактном классе сигналов). Пусть $K \subset H^n$ — компактный класс сигналов и $f \in C(K)$ — непрерывный скалярный функционал (“критерий/оценка”). Тогда для любого $\varepsilon > 0$ существует функция вида

$$g(x_1, \dots, x_n) = \sum_{j=1}^{r} \zeta_j \Big(\sum_{i=1}^{n} \varphi_{ji}(x_i) \Big),$$

где $\varphi_{ji} \in H^*$ (линейные измерения/фильтровые каналы) и $\zeta_j \colon \mathbb{R} \to \mathbb{R}$ непрерывны (скалярные нелинейности), такая что

$$\sup_{(x_1,\dots,x_n)\in K} |f(x_1, \dots, x_n) - g(x_1, \dots, x_n)| < \varepsilon.$$

Доказательство. Утверждение следует непосредственно из Теоремы 1 в [3]; ниже приводим только перевод нотации к принятой в статье форме.

Положим $X := H^n$ и рассматриваем X как гильбертово пространство с нормой

$$\|(x_1, \dots, x_n)\|^2 := \sum_{i=1}^{n} \|x_i\|_H^2.$$

По Теореме 1 из [3] для любого компакта $K \subset X$ и любого $f \in C(K)$ существует аппроксимация g на K с точностью ε конечной суммой

$$g(x) = \sum_{j=1}^{r} \zeta_j\big(\Phi_j(x)\big),$$

где $\Phi_j \in X^*$ — непрерывные линейные функционалы, а $\zeta_j \colon \mathbb{R} \to \mathbb{R}$ — непрерывные функции.

Разложим каждый $\Phi_j \in X^* = (H^n)^*$ по координатам. Определим $\varphi_{ji} \in H^*$ формулой

$$\varphi_{ji}(u) := \Phi_j(0, \dots, 0, u, 0, \dots, 0),$$

где u стоит на i-й позиции. Тогда по линейности для любого $x = (x_1, \dots, x_n) \in H^n$ имеем

$$\Phi_j(x_1, \dots, x_n) = \sum_{i=1}^{n} \varphi_{ji}(x_i),$$

и получаем требуемую форму

$$g(x_1, \dots, x_n) = \sum_{j=1}^{r} \zeta_j\Big(\sum_{i=1}^{n} \varphi_{ji}(x_i)\Big).$$

Теорема 2 (векторнозначный случай). Пусть $K \subset H^n$ — компактный класс сигналов, Y — банахово пространство и $f \in C(K, Y)$ — непрерывное отображение. Тогда для любого $\varepsilon > 0$ существует аппроксимация *конечного ранга* вида

$$g(x_1, \dots, x_n) = \sum_{j=1}^{r} y_j\, \zeta_j\Big(\sum_{i=1}^{n} \varphi_{ji}(x_i)\Big), \qquad y_j \in Y,$$

где $\varphi_{ji} \in H^*$ — непрерывные линейные функционалы, а $\zeta_j \colon \mathbb{R} \to \mathbb{R}$ — непрерывные скалярные функции, такая что

$$\sup_{(x_1,\dots,x_n)\in K} \|f(x_1, \dots, x_n) - g(x_1, \dots, x_n)\|_Y < \varepsilon.$$

3 Практические приложения и интерпретация в терминах ЦОС

Ниже приведены типовые сценарии, в которых предпосылка компактности класса сигналов и представление через конечное число линейных измерений естественны, а также обсуждается инженерная интерпретация построения.

3.1 Примеры компактных классов сигналов $\boldsymbol{K}$

Формально, основной результат носит вид "на любом компакте $K \subset H^n$"; на практике компактность часто обеспечивается совокупностью (i) ограниченности энергии и (ii) некоторой регулярности или параметричности модели. Ниже перечислены стандартные конструкции.

(a) Полосоограниченные сигналы с ограниченной энергией. Пусть $H = L^2([0, T])$ (или $L^2(\mathbb{R})$ после оконного ограничения), и рассмотрим множество сигналов, спектр которых лежит в фиксированной полосе, а норма ограничена:

$$K = \{x \in H : \operatorname{supp} \widehat{x} \subset [-\Omega, \Omega], \|x\|_2 \leq R\}.$$

В подходящих топологиях (например, при замене L^2-нормы на более сильную, либо при явном введении гладкости/ограничения производных) такие множества ведут себя как "почти конечномерные" и допускают аппроксимацию конечным числом коэффициентов.

(b) Классы с ограниченной гладкостью (Соболевские шары). Если $H = H^s([0, T])$ и $K = \{x : \|x\|_{H^s} \leq R\}$, то вложение $H^s \hookrightarrow H$ при $s > 0$ даёт компактность образа шара в более слабой норме (типичный способ получить компакт $K \subset H$ как "энергия + гладкость").

(c) Параметрические семейства конечной размерности. Если сигналы задаются конечным числом параметров $\theta \in \Theta \subset \mathbb{R}^p$ (например, сумма затухающих синусоид, либо шаблон с неизвестной задержкой/-масштабом), то $K = \{x(\theta) : \theta \in \Theta\}$ будет компактным при компактной Θ и непрерывности отображения $\theta \mapsto x(\theta)$.

(d) Классы "событий" в сейсмике/аудио. Практически часто фиксируют: ограничение длительности/окна, ограничение полосы, ограничение динамического диапазона и минимальную гладкость; это эмпирически стабилизирует класс сигналов и приближает его к компактному множеству в выбранной норме.

3.2 Линейные измерения как банк фильтров

Линейные функционалы $\varphi \in H^*$ удобно интерпретировать как "каналы" анализа. Если H — гильбертово пространство, то по теореме Рисса для каждого $\varphi \in H^*$ существует вектор $h \in H$ такой, что $\varphi(x) = \langle x, h \rangle$ (корреляция/согласованная фильтрация). Для многоканального входа $x = (x_1, \ldots, x_n)$ величины $\varphi_{ji}(x_i)$ можно понимать как выходы банка фильтров, применённых к каждому каналу, с последующим суммированием по i на уровне "признака" j. В задачах телеметрии и событийных сигналов часто встречается неравномерная сетка наблюдений; тогда удобно рассматривать H как пространство функций/временных рядов на такой сетке (с подходящим скалярным произведением), а линейные измерения φ_{ji} интерпретировать как обобщённые свёртки/взвешенные суммы по доступным отсчётам. На практике сами φ_{ji} и параметры нелинейностей ζ_j часто выбирают из параметризованных семейств и обучают по данным градиентными методами; Теоремы 1–2 говорят, что существует достаточно богатый класс таких параметризаций, обеспечивающий универсальность аппроксимации на компактном K.

3.3 Нелинейности ζ_j и связь с типовыми ЦОС-цепочками обработки

Скалярные нелинейности $\zeta_j\colon \mathbb{R} \to \mathbb{R}$ моделируют статические преобразования каналов (огибающая, энергия, компрессия, пороги), после чего выполняется агрегация:

$$g(x) = \sum_{j=1}^{r} \zeta_j\Big(\sum_{i=1}^{n} \varphi_{ji}(x_i)\Big).$$

Это соответствует шаблону "linear filterbank + pointwise nonlinearity + pooling/aggregation".

Связь с современными архитектурами и практическими приложениями. Схема "линейные измерения + скалярные нелинейности + агрегация" тесно связана с практикой построения признаков в DSP и с современными моделями для функций и временных рядов: банки фильтров и последующие энергии/огибающие лежат в основе многих аудио- и сейсмических пайплайнов, а также 1D-CNN и родственных архитектур. С точки зрения обучения операторов и карт между функциональными пространствами полезно сопоставление с нейрооператорами (например, DeepONet) [4], классическими результатами универсальной аппроксимации в конечномерном случае [1, 2, 5] и оценками сложности в высокоразмерной вероятности [6].

3.4 Примеры: детектирование и оценка

Детектирование события. Целевая функция $f(x)$ может быть, например, оценкой наличия события в окне (или калиброванной оценкой); результат утверждает аппроксимируемость такой непрерывной оценки через конечный набор фильтровых каналов и статические нелинейности.

Оценка параметра. Для параметров вроде задержки/частоты часто строят скалярный критерий качества, который затем оптимизируют по параметру; сам критерий (для фиксированного кандидата) подпадает под рассматриваемый класс непрерывных функционалов.

Классификационная оценка. Во многих задачах достаточно вещественной "оценки" до применения порога/softmax; на компактном K она аппроксимируема моделью "линейные признаки + сумма скалярных нелинейностей". Кроме того, теорема применима к непрерывным функционалам в задачах восстановления сигналов/изображений, таким как “качество реконструкции” или регуляризатор (штраф), используемый при оптимизации.

3.5 Векторнозначный случай

Векторнозначный вариант (Теорема 2) далее используется в инженерных интерпретациях (например, для векторных выходов моделей). Аккуратную формулировку и обсуждение см. в Теореме 2 (§2); ниже сосредоточимся на интерпретациях и примерах из ЦОС, не повторяя определение.

3.6 Пример: ЭКГ/ЭЭГ и проекции на функции Эрмита и Лагерра

Рассмотрим одноканальный сигнал $x \in H$ (например, окно ЭКГ или ЭЭГ длины T), либо многоканальный сигнал $x = (x_1, \dots, x_n) \in H^n$ (многоканальная ЭЭГ). В рамках основной теоремы (скалярный выход) мы аппроксимируем *целевой функционал*

$$f\colon K \to \mathbb{R},$$

где $K \subset H^n$ — компактный класс фрагментов сигналов, а значение $f(x)$ есть интересующая нас непрерывная характеристика. В медицинской обработке сигналов типичные примеры таких функционалов:

- $f(x)$ — *оценка детекции* события (например, QRS-комплекс в ЭКГ; эпилептиформная активность в ЭЭГ) в окне;
- $f(x)$ — *оценка качества* записи (артефакты/шум/плохой контакт);
- $f(x)$ — *регрессионная оценка* параметра (например, ширина QRS или интервал QT при фиксированной процедуре сегментации).

$$\boxed{\begin{array}{c}\textbf{сигнал } x \;\longrightarrow\; \textbf{банк проекций (Эрмит/Лагерр): } \{\varphi_j(x)\}_{j=1}^r \;\longrightarrow\\ \textbf{нелинейности: } \{\zeta_j(\varphi_j(x))\}_{j=1}^r \;\longrightarrow\; \textbf{суммирование/выход:}\\ g(x) = \sum_{j=1}^r \zeta_j(\varphi_j(x))\end{array}}$$

Рис. 1: Блок-схема универсального аппроксиматора: сигнал → банк проекций (Эрмит/Лагерр) → скалярные нелинейности → агрегация.

Далее будем придерживаться именно этой инженерной интерпретации (рис. 1).

Где здесь линейные измерения (банк проекций). В качестве линейных функционалов $\varphi_j \in H^*$ (или $\varphi_{ji} \in H^*$ в многоканальном случае) удобно взять проекции на фиксированную систему базисных функций:

- *функции Эрмита* (локализованные “гауссовы” формы, удобные для морфологии пиков и их производных);

- *функции Лагерра* (экспоненциально затухающие функции, естественные для каузальных моделей и систем с памятью).

Тогда измерения имеют вид $\varphi_j(x) = \langle x, \psi_j \rangle$ (или $\varphi_{ji}(x_i) = \langle x_i, \psi_{ji} \rangle$), где ψ_j выбираются из соответствующих семейств Эрмита/Лагерра.

Нормировка и практические детали. Для окна длины T обычно вводят (i) центрирование по времени (выравнивание относительно предполагаемого события, например R-пика), (ii) масштабирование времени $t \mapsto (t-t_0)/\sigma$ под типичную ширину комплекса (или под доминирующую полосу в ЭЭГ), и (iii) предварительную фильтрацию/удаление дрейфа. Тогда "эрмитовы" проекции естественно писать как

$$\varphi_k(x) = \int_0^T x(t)\, h_k\!\left(\frac{t-t_0}{\sigma}\right) dt,$$

где h_k — ортонормированные функции Эрмита (гаусс $\times$ полином Эрмита), а параметры t_0, σ фиксируются заранее или выбираются по простой процедуре выравнивания. Аналогично, "лагерровы" проекции для каузальной параметризации можно записать как

$$\varphi_k(x) = \int_0^T x(t)\, \ell_k(t; \alpha)\, dt,$$

где $\ell_k(\cdot; \alpha)$ — ортонормированные функции Лагерра с параметром затухания $\alpha > 0$ (задаёт "память" и масштаб времени). В дискретном времени интегралы заменяются скалярным произведением $\langle x, \psi_k \rangle = \sum_{m=1}^{M} x[m]\psi_k[m]$.

Почему Эрмит удобен для ЭКГ, а Лагерр — для моделей с памятью. Функции Эрмита локализованы и хорошо описывают "пикообразную" морфологию (QRS и его производные) небольшим числом коэффициентов; это даёт компактный набор линейных признаков, устойчивых к шуму при корректной нормировке. Функции Лагерра, напротив, тесно связаны с каузальными моделями и аппроксимацией линейных динамических систем с затухающей памятью: проекции на базис Лагерра можно интерпретировать как выходы фиксированного банка рекурсивных фильтров (что удобно в онлайновой обработке) [7, 8].

Где здесь нелинейности и агрегирование. Теорема утверждает, что при подходящем выборе числа каналов r и непрерывных скалярных нелинейностей $\zeta_j \colon \mathbb{R} \to \mathbb{R}$ функционал f на K можно равномерно приблизить выражением

$$g(x) = \sum_{j=1}^{r} \zeta_j\big(\varphi_j(x)\big) \quad \text{(одноканальный случай)},$$

или

$$g(x_1, \dots, x_n) = \sum_{j=1}^{r} \zeta_j \Big(\sum_{i=1}^{n} \varphi_{ji}(x_i) \Big) \quad \text{(многоканальный случай).}$$

Здесь (i) проекции на Эрмита/Лагерра играют роль конечного банка измерений/признаков, (ii) ζ_j моделируют точечные нелинейности (энергия, модуль, порог/компрессия), а (iii) суммирование по j соответствует pooling/агрегации в итоговую оценку $g(x) \approx f(x)$.

Типовые нелинейности и связь с привычными признаками. Например, выбор $\zeta_j(u) = u^2$ соответствует агрегированию энергий по подпространствам (аналогично энергии в подполосах), $\zeta_j(u) = |u|$ даёт робастную "амплитудную" агрегацию, а насыщаемые/пороговые ζ_j моделируют детекторные эвристики. В многоканальном ЭЭГ случае внутренняя сумма $\sum_i \varphi_{ji}(x_i)$ позволяет (при подходящем выборе φ_{ji}) реализовать пространственные фильтры (например, фиксированные линейные комбинации каналов) до скалярной нелинейности.

Замечание про компактность класса K (что считать реалистичным). Чтобы интерпретация "универсальности" была содержательной, удобно явно моделировать K как класс окон с ограниченной энергией и гладкостью/полосой (и, для ЭКГ, с ограниченной вариабельностью длительностей и амплитуд после выравнивания). Тогда конечномерная проекция на первые m функций Эрмита/Лагерра действительно даёт малую аппроксимационную ошибку на K, и число каналов можно понимать как "сложность морфологического класса".

Указания на литературу (Эрмит/Лагерр в ЭКГ/ЭЭГ). Классические примеры разложений ЭКГ в базисе функций Эрмита (в том числе для компактного описания морфологии QRS) см. [9, 10]. Применение обобщённых функций Лагерра (Laguerre expansions) и связанных ортонормированных базисов в задачах моделирования динамики/систем с памятью (что часто используют как "каузальный" банк проекций) обсуждается, например, в [7, 8]. Для современных обзоров по автоматическому анализу ЭЭГ/ЭКГ, где ортогональные базисы/спектрально-временные разложения фигурируют как стадия извлечения признаков, см. [11, 12].

Векторнозначный случай. Теорема 2 формально покрывает и векторные выходы; ниже приведём только инженерные интерпретации и примеры (без повторения формулы аппроксиматора).

Примеры из ЦОС.

- **Многоклассовая классификация (вектор оценок/логарифмов отношения шансов).** Пусть $Y = \mathbb{R}^C$, а $f(x)$ возвращает вектор непрерывных оценок/логарифмов отношения шансов по C классам для сигналов из компактного K. Теорема 2 даёт аппроксимацию, в которой один и тот же банк линейных измерений порождает скалярные “коэффициенты” $\zeta_j(\cdot)$, а векторы $y_j \in \mathbb{R}^C$ играют роль выходных “атомов” (базисных направлений в пространстве классов).

- **Оценка спектральной огибающей или подполосных энергий.** Пусть $Y = \mathbb{R}^m$, а $f(x)$ возвращает, например, m-мерный вектор энергий по подполосам (mel/erb) или параметров огибающей на фиксированном окне. На компактном классе входов K (энергия+гладкость/полоса) такая карта аппроксимируема конечным рангом; в DSP-терминах это обосновывает “общий” анализ (фильтры/проекции) с последующим вычислением набора выходных характеристик.
- **Карта признаков во времени (дискретизированная траектория).** Если интересует выход $f(x)$ как временной профиль некоторой величины на окне (например, локальная энергия/детектор), можно рассматривать $Y = \mathbb{R}^T$ после дискретизации по времени. Тогда Теорема 2 утверждает, что траектория аппроксимируется суммой фиксированных временных шаблонов $y_j \in \mathbb{R}^T$ с коэффициентами, зависящими от линейных измерений входа.
- **Многоканальная оценка направления прихода/параметров массива.** Для задач с массивом датчиков выходом часто является вектор параметров/оценок по сетке направлений. Это снова $Y = \mathbb{R}^m$, и конечный ранг соответствует разложению по ограниченному набору “пространственных” шаблонов y_j при коэффициентах, получаемых из линейных измерений сигналов каналов.

3.7 Число каналов и сложность класса

Инженерная интерпретация: чем меньше ε-энтропия K, тем меньшим числом линейных “каналов измерения” можно аппроксимировать любой непрерывный функционал на K (см. классические оценки ε-энтропии у Колмогорова–Тихомирова [13]). Напомним, что (логарифмическая) ε-энтропия может быть определена как

$$H_\varepsilon(K) := \log N_\varepsilon(K),$$

где $N_\varepsilon(K)$ — минимальное число ε-шаров (в норме пространства сигналов), покрывающих K. Интуитивно, $N_\varepsilon(K)$ измеряет “сколько эффективных степеней свободы” остаётся у сигналов при требуемой точности ε.

Для многих “физически ограниченных” классов сигналов $H_\varepsilon(K)$ растёт при $\varepsilon \downarrow 0$ полиномиально по $1/\varepsilon$. Например, для классов с ограниченной гладкостью (типичный мысленный образ: шар в пространстве Соболева на отрезке) качественно имеет место рост вида

$$H_\varepsilon(K) \asymp \varepsilon^{-1/s} \qquad (\varepsilon \to 0),$$

где $s > 0$ — параметр гладкости (чем больше s, тем "проще" класс и тем медленнее рост энтропии). В многомерных/многоканальных задачах аналогичные оценки обычно содержат "эффективную размерность" (например, множитель n и/или степень, зависящую от размерности области определения).

С практической точки зрения это даёт полезное "ощущение масштаба" числа каналов: чтобы гарантировать аппроксимацию на K с точностью порядка ε, требуется число независимых измерений/коэффициентов, которое в худшем случае растёт примерно как $N_\varepsilon(K)$ (или как $H_\varepsilon(K)$ с точностью до логарифмических факторов), то есть для гладких классов — существенно медленнее, чем экспоненциально.

4 Связанные идеи и близкие формулировки в литературе (ЦОС и смежные области)

Ниже перечислены направления, близкие по *архитектурной идее* (фильтрбанк/линейные измерения + поэлементные нелинейности + агрегация), а также по общей логике "ограниченность класса сигналов $\Rightarrow$ конечномерный заместитель". Для каждого направления даём короткое описание и ссылки на источники, оформленные в списке литературы.

4.1 Инвариантность и устойчивость представлений ("рассеяние" и свёрточные нейронные сети как извлечение признаков)

Преобразования рассеяния (scattering) строят представления каскадом "фильтрбанк $\to$ модуль $\to$ усреднение" и дают строгие утверждения об инвариантности и устойчивости к деформациям; см. [14, 15]. Близкие по духу результаты о стабильности/липшицевости и извлечении признаков для более общих каскадов типа свёрточных нейронных сетей (CNN) обсуждаются в [16].

4.2 Нелинейные системы с памятью и "универсальные" разложения (Volterra/Wiener, затухающая память)

В системной идентификации и классической ЦОС широко используются Volterra/Wiener-подобные представления как универсальные модели нелинейных систем с памятью при дополнительных предпосылках (например, затухающая память (fading memory)); см. [17, 18].

4.3 Аппроксимация векторозначных отображений конечным рангом

Для непрерывных отображений $f\colon K \to Y$ на компактах (где Y — банахово пространство) изучается аппроксимация функциями с конечномерным образом (finite-dimensional range/finite-rank approximations); см., например, [19].

4.4 EMD как конечномерное представление

Эмпирическая модовая декомпозиция (empirical mode decomposition, EMD) широко используется для анализа нестационарных сигналов (в том числе ЭКГ/ЭЭГ) и представляет вход x как сумму "внутренних мод" (intrinsic mode functions, IMF) и остатка. Алгоритм EMD является адаптивным и нелинейным, однако с прикладной точки зрения его часто используют как способ получить конечное число компонент, после чего вычисляют скалярные или векторные характеристики (энергии мод, показатели нерегулярности, мгновенные частоты и т.п.) и агрегируют их в непрерывную оценку/логарифм отношения шансов. В контексте настоящей статьи EMD можно рассматривать как практический механизм построения конечномерного описания сигналов из некоторого "реалистичного" класса K. После выбора конечного набора измерений/компонент последующая стадия "поэлементные нелинейности $\to$ агрегация" естественным образом укладывается в рассматриваемую схему аппроксимации непрерывных функционалов на компактах. Базовые ссылки на EMD и примеры применения к ЭКГ/ЭЭГ см. [20–22].

4.4.1 Пример интерпретации EMD (модифицированного) для анализа ЭЭГ

В [23] анализ ЭЭГ строится как "(модифицированный) EMD $\to$ признаки $\to$ непрерывная оценка/логарифм отношения шансов": сигнал разлагается на конечное число мод (IMF) и остаток, затем по модам вычисляются числовые характеристики (энергии/спектральные показатели/меры нерегулярности и т.п.), которые используются для классификации/оценивания.

С точки зрения настоящей статьи, этот подход можно трактовать как практический способ получить конечномерное описание $F(x) \in \mathbb{R}^p$ для "реалистичного" класса сигналов K (за счёт адаптивной декомпозиции), а затем аппроксимировать целевой функционал на K. Если K компактен, то любая непрерывная цель $f|_K$ допускает равномерное приближение архитектурой

$$g(x_1, \dots, x_n) = \sum_{j=1}^{r} \zeta_j \Big(\sum_{i=1}^{n} \varphi_{ji}(x_i) \Big),$$

то есть конечным числом линейных измерений и скалярных нелинейностей. В этой интерпретации модификации EMD естественно понимать как стабилизацию/согласование признаков на K, уменьшающую требуемую сложность последующей аппроксимации.

5 Заключение

Мы показали, что на компактных классах многоканальных сигналов $K \subset H^n$ любая непрерывная целевая зависимость может быть равномерно аппроксимирована архитектурой вида "конечный банк линейных измерений $\to$ скалярные нелинейности $\to$ агрегация", а в векторозначном случае — аппроксимацией конечного ранга в пространстве выходов. В инженерных терминах это даёт строгое обоснование широко используемого шаблона построения признаков и моделей в ЦОС: фильтрбанк/проекции, поэлементные нелинейности (энергия, модуль, компрессия) и pooling/суммирование.

Ключевой предпосылкой является компактность класса сигналов K: именно она позволяет свести задачу аппроксимации к конечномерному "заместителю" (существует конечномерная проекция, равномерно приближающая элементы K), после чего применима классическая универсальная аппроксимация в конечномерном пространстве. Если компактность нарушается (например, допускаются сигналы с неограниченной энергией, произвольной длительностью окна, неограниченной полосой/гладкостью или без какого-либо контроля "размерности" класса), то равномерная аппроксимация на всём множестве уже, вообще говоря, не гарантируется: могут появляться последовательности сигналов, для которых никакой фиксированный конечный набор линейных измерений не обеспечивает одновременный контроль ошибки. Практически это означает, что универсальность следует интерпретировать как утверждение *на ограниченном, физически реалистичном классе* входов (энергия, длительность, полоса/регулярность, нормировка), а не как глобальную гарантию для произвольных сигналов из бесконечномерного пространства.

Новизна по сравнению с направлениями из раздела "Связанные идеи" состоит в следующем. Во-первых, линии рассеяния (scattering) и близкие теории признаков фокусируются преимущественно на *устойчивости* и *инвариантности* конкретных каскадов (часто при фиксированном выборе нелинейности), тогда как в нашей работе центральным является утверждение об *универсальности аппроксимации* на произвольных компактах при гибком выборе линейных измерений и скалярных нелинейностей. Во-вторых, классические Volterra/Wiener-представления дают универсальные модели для операторов при дополнительных структурных предпосылках (например, *fading memory*) и описывают нелинейность через многомерные ядра/свёртки; наш результат, напротив, выделяет класс аппроксиматоров, реализуемых конечным числом *линейных* *функционалов* и *скалярных* нелинейностей, что ближе к современной парадигме "измерения/признаки + нелинейный микшер". В-третьих, результаты по аппроксимации векторозначных непрерывных отображений в функциональном анализе описывают существование конечномерного приближения, но обычно не дают явной интерпретации в терминах банка измерений сигналов; в нашей постановке такая интерпретация является частью самой формулы аппроксиматора.

В целом работа проясняет различие между двумя тезисами — "устойчивые и инвариантные признаки" и "достаточная выразительная мощность" — и показывает, что на компактных классах сигналов архитектуры семейства "filterbank + pointwise nonlinearity + pooling" обладают универсальностью по отношению к непрерывным целевым зависимостям. Это может служить теоретической базой для дальнейших оценок сложности, построения адаптивных банков измерений и анализа компромиссов между устойчивостью и выразительностью.

Авторские декларации

Финансирование

Работа выполнена при поддержке госбюджетной темы НИР ВМК МГУ.

Доступность данных и программного кода

Не предоставляются.

Конфликт интересов

Отсутствует.

Вклад авторов

А. С. Крылов — разработка математических методов, написание текста статьи.

Список литературы

[1] *Cybenko George.* Approximation by superpositions of a sigmoidal function // *Mathematics of Control, Signals, and Systems.* — 1989. — Vol. 2. — Pp. 303–314.

[2] Multilayer feedforward networks with a nonpolynomial activation function can approximate any function / Moshe Leshno, Vladimir Ya. Lin, Allan Pinkus, Shimon Schocken // *Neural Networks.* — 1993. — Vol. 6. — Pp. 861–867.

[3] *Krylov A., Penkin M.* Universal Approximation of Continuous Functionals on Compact Subsets via Linear Measurements and Scalar Nonlinearities // *arXiv preprint arXiv:2602.03290.* — 2026.

[4] *Lu Lu, Jin Pengzhan, Karniadakis George Em.* Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators // *Nature Machine Intelligence.* — 2021. — Vol. 3. — Pp. 218–229.

[5] *Barron Andrew R.* Universal approximation bounds for superpositions of a sigmoidal function // *IEEE Transactions on Information Theory.* — 1993. — Vol. 39. — Pp. 930–945.

[6] *Vershynin Roman.* High-Dimensional Probability: An Introduction with Applications in Data Science. — Cambridge University Press, 2018.

[7] *Wahlberg Bo.* System identification using Laguerre models // *IEEE Transactions on Automatic Control.* — 1991. — Vol. 36. — Pp. 551–562.

[8] *Heuberger P. S. C., Van den Hof P. M. J., Bosgra O. H.* Modelling and Identification with Rational Orthogonal Basis Functions. — Springer, 2005.

[9] Clustering ECG complexes using Hermite functions and self-organizing maps / M. Lagerholm, C. Peterson, G. Braccini et al. // *IEEE Transactions on Biomedical Engineering.* — 2000. — Vol. 47. — Pp. 838–848.

[10] *Addison Paul S.* Wavelet Transforms and the ECG: A Review. — Institute of Physics Publishing, 2005.

[11] Wavelet-based EEG processing for computer-aided seizure detection and epilepsy diagnosis / Oliver Faust, U. Rajendra Acharya, Hojjat Adeli, Amir Adeli // *Seizure.* — 2015. — Vol. 26. — Pp. 56–64.

[12] A deep convolutional neural network model to classify heartbeats / U. Rajendra Acharya, Shu Lih Oh, Yuki Hagiwara et al. // *Computers in Biology and Medicine.* — 2017. — Vol. 89. — Pp. 389–396.

[13] *Kolmogorov A. N., Tikhomirov V. M.* Entropy and capacity of sets in functional spaces // *Uspekhi Mat. Nauk.* — 1959. — Vol. 14, no. 2. — Pp. 3–86. — English transl.: Russian Mathematical Surveys.

[14] *Mallat Stéphane.* A Wavelet Tour of Signal Processing. — 2 edition. — Academic Press, 1999.

[15] *Bruna Joan, Mallat Stéphane.* Invariant scattering convolution networks // *IEEE Transactions on Pattern Analysis and Machine Intelligence.* — 2013. — Vol. 35. — Pp. 1872–1886.

[16] *Wiatowski Michal, Bölcskei Helmut.* A mathematical theory of deep convolutional neural networks for feature extraction // *IEEE Transactions on Information Theory.* — 2018. — Vol. 64. — Pp. 1845–1866.

[17] *Boyd Stephen, Chua Leon O.* Fading memory and the problem of approximating nonlinear operators with Volterra series // *IEEE Transactions on Circuits and Systems.* — 1985. — Vol. 32. — Pp. 1150–1161.

[18] *Wiener Norbert.* Nonlinear Problems in Random Theory. — MIT Press, 1958.

[19] *Shuchat A. H.* Approximation of vector-valued continuous functions // *Proceedings of the American Mathematical Society.* — 1972. — Vol. 31. — Pp. 97–103.

[20] The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis / Norden E. Huang, Zheng Shen, Steven R. Long et al. // *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences.* — 1998. — Vol. 454. — Pp. 903–995.

[21] *Wu Zhaohua, Huang Norden E.* Ensemble empirical mode decomposition: a noise-assisted data analysis method // *Advances in Adaptive Data Analysis.* — 2009. — Vol. 1, no. 1. — Pp. 1–41.

[22] EEGLAB, SIFT, NFT, BCILAB, and ERICA: new tools for advanced EEG processing / Arnaud Delorme, Tim Mullen, Christian Kothe et al. // *Computational Intelligence and Neuroscience.* — 2011. — P. 130714.

[23] *Богданова О. А., Пономарев Т. Д., Крылов А. С.* Анализ ЭЭГ на основе модифицированного метода эмпирических мод // *Прикладная математика и информатика: Труды факультета ВМК МГУ имени М. В. Ломоносова.* — 2025. — Vol. 78. — Pp. 28–36. — Москва: ООО МАКС Пресс.