ДОКЛАДЫ РОССИЙСКОЙ АКАДЕМИИ НАУК. МАТЕМАТИКА, ИНФОРМАТИКА, ПРОЦЕССЫ УПРАВЛЕНИЯ, 2023, том 514, № 2, с. 28–38

УДК 519.6

ОПТИМИЗАЦИЯ ФИЗИКО-ИНФОРМИРОВАННЫХ НЕЙРОННЫХ СЕТЕЙ ДЛЯ РЕШЕНИЯ НЕЛИНЕЙНОГО УРАВНЕНИЯ ШРЕДИНГЕРА

© 2023 г. И. А. Чупров[1,*], Ц. Гао[1,], Д. С. Ефременко[1], Е. А. Казаков[1], Ф. А. Бузаев[1], В. В. Земляков[1]**

Представлено академиком РАН А.Л. Семеновым
Поступило 01.09.2023 г.
После доработки 15.09.2023 г.
Принято к публикации 15.09.2023 г.

Физико-информированные нейронные сети (Physics Informed Neural Networks – PINN) являются перспективным методом решения уравнений в частных производных с помощью машинного обучения. В работе рассмотрено применение PINN к нелинейному уравнению Шредингера для описания распространения сигнала в оптическом волокне. Исследуются факторы, определяющие сходимость PINN с физической точки зрения. Получены оценки области сходимости метода по длине волокна и энергии импульса. Показано, что применение синусоидальной активационной функции, а также весов для слагаемых функции потерь позволяет увеличить область сходимости PINN относительно длины волокна и энергии импульса. Произведено обобщение метода (мета-PINN), позволяющее решать уравнение при различных его параметрах с помощью предобученной нейронной сети.

Ключевые слова: физико-информированные нейронные сети, нелинейное уравнение Шредингера, нелинейная волоконная оптика, тонкая настройка нейронных сетей

DOI: 10.31857/S2686954323601586, **EDN:** GOJIHT

1. ВВЕДЕНИЕ

Одним из перспективных направлений применения машинного обучения в научных вычислениях являются нейронные сети с учетом физических законов (Physics Informed Neural Networks – PINN). PINN относятся к универсальным аппроксиматорам функций, которые обучаются с учетом физических законов [1]. PINN часто рассматривают в качестве альтернативы традиционным методам решения уравнений в частных производных (Partial Differential Equations – PDE) [2]. Методология PINN используется во многих областях вычислительной физики. В данной работе рассмотрено применение PINN к нелинейному уравнению Шредингера, которое описывает распространение сигнала в оптическом волокне [3]. Ранее PINN применялся к решению данного типа уравнений, в частности [4]. В работе [5] проведено сравнение PINN с методом разделения шагов. В работе [6] PINN использовался для расчета много-модового оптического волокна. К преимуществам PINN относят следующие его особенности:

1. Методология PINN универсальна и может быть использована для произвольного уравнения в частных производных с минимальными изменениями в коде [7].

2. Теоретически PINN может обойти по скорости решения метод разделения шагов при достаточно большой длине оптоволокна [8].

3. Обучение PINN может быть легко распараллелено [9].

4. Благодаря использованию предобученных PINN-моделей время обучения для нового сценария может быть существенно сокращено [2].

Однако у метода PINN есть недостатки:

1. PINN сходится не всегда из-за сложности функции потерь [10]; может возникнуть проблема исчезающего градиента [11] в процессе обучения.

2. Оценка его точности без референсного решения затруднительна, не решены теоретические вопросы сходимости метода [12].

3. Время обучения PINN может превышать время решения уравнения традиционным методом (например, методом конечных элементов) [13].

В отличие от многочисленных работ, посвященных применению PINN (в которых, в основном, говорится о том, что PINN способны решать

[1]*Московский исследовательский центр Huawei (Huawei Moscow Research Center), Москва, Россия*
**E-mail: chuprov.ivan@huawei.com*
***E-mail: gaojiexing@huawei.com*

заданные PDE), не хватает публикаций, в которых бы исследовалась область применимости PINN или освещались факторы, определяющие эффективность PINN. В связи с этим часто остается неясным, следует ли использовать PINN вместо традиционных решателей PDE, а если следует, то какая конфигурация нейронной сети обеспечивает наилучшую эффективность PINN. Отметим, что в публикациях о NLSE [4, 5, 8] конфигурация PINN и количество точек дискретизации постулировались, а вопросы сходимости подробно не рассматривались.

В данной работе целью является определение зависимости сходимости PINN от параметров нелинейного уравнения Шредингера. Также рассматриваются методы, позволяющие расширить область применения PINN. Итоговая наша цель – сформулировать практические рекомендации о применении PINN к рассматриваемой задаче.

2. ТЕОРИЯ

2.1. Нелинейное уравнение Шредингера

Распространение лазерного импульса в оптоволоконном кабеле в системе координат, движущейся с групповой скоростью импульса, описывается нелинейным уравнением Шредингера [3]:

$$\frac{\partial A}{\partial z} + \frac{i\beta_2}{2}\frac{\partial^2 A}{\partial T^2} + \frac{\alpha}{2}A = i\gamma|A|^2 A, \tag{1}$$

где $A(z,T)$ – комплексная временная огибающая электрического поля, z – пространственная координата в оптоволокне, β_2 – коэффициент дисперсии второго порядка, T – временная развертка импульса, α – коэффициент поглощения, а γ – коэффициент Керра. Заметим, что $A = u + iv$ является комплексной функцией, и поэтому уравнение (1) является системой из двух уравнений. Учитываются три физических эффекта, а именно затухание, дисперсия и самофазная модуляция через члены при α, β_2 и γ соответственно. Каждый из этих эффектов проявляет себя на расстоянии, большем определенного, называемого характеристической длиной. Можно ввести характеристическую длину затухания L_a,

$$L_a = \frac{1}{\alpha}. \tag{2}$$

дисперсионную длину L_d

$$L_d = \frac{T_0^2}{|\beta_2|}, \tag{3}$$

и нелинейную длину L_n

$$L_n = \frac{1}{\gamma P_0}, \tag{4}$$

где T_0 и P_0 – начальные ширина и мощность в пике пульса соответственно. По соотношению этих длин можно оценить, насколько сильно будет проявляет себя соответствующий эффект. Например, если $L_n \gg L_d$ и $z \sim L_d$, то нелинейный эффект будет пренебрежимо мал по сравнению с дисперсионным эффектом. В данной работе без потери общности ограничимся анализом гауссовского импульса в качестве граничных условий:

$$A(0,T) = A_0(T), \tag{5}$$

где

$$A_0(T) \sim \exp\left\{-\frac{T^2}{2T_0^2}\right\}, \tag{6}$$

импульс на входе оптоволокна, T_0 – полуширина импульса, а $[-T_{\max}, T_{\max}]$ – временное окно. Начальные условия имеют вид

$$A(z, \pm T_{\max}) = 0. \tag{7}$$

2.2. Замена переменных

При решении уравнения (1) часто прибегают к преобразованию системы координат [14]:

$$z = k_1 L_d \xi, \tag{8}$$

$$T = k_2 T_0 t, \tag{9}$$

где

$$k_1 = \frac{L_{\max}}{L_d} \tag{10}$$

и

$$k_2 = \frac{T_{\max}}{T_0}. \tag{11}$$

Также A нормируется на пиковую мощность P_0:

$$A = \sqrt{P_0}U. \tag{12}$$

Уравнение для нормированного сигнала U в преобразованной системе координат (ξ, t) имеет вид:

$$\frac{\partial U}{\partial \xi} + b\frac{\partial^2 U}{\partial T^2} + aU = ic|U|^2 U, \tag{13}$$

где

$$a = \frac{\alpha L_d k_1}{2}, \quad b = \frac{k_1\beta_2}{k_2^2}, \quad c = \frac{k_1 L_d}{L_n}. \tag{14}$$

В данной работе расчеты проводятся для длины волны $\lambda = 1030$ нм и коэффициента $\beta_2 = 0.019164$. Параметр поглощения α принят равным нулю. Временной интервал $T_{\max}$ составляет 100 пс. Параметр Керра γ равен 0.0011465. Энергия импульса и длина волокна варьирует-

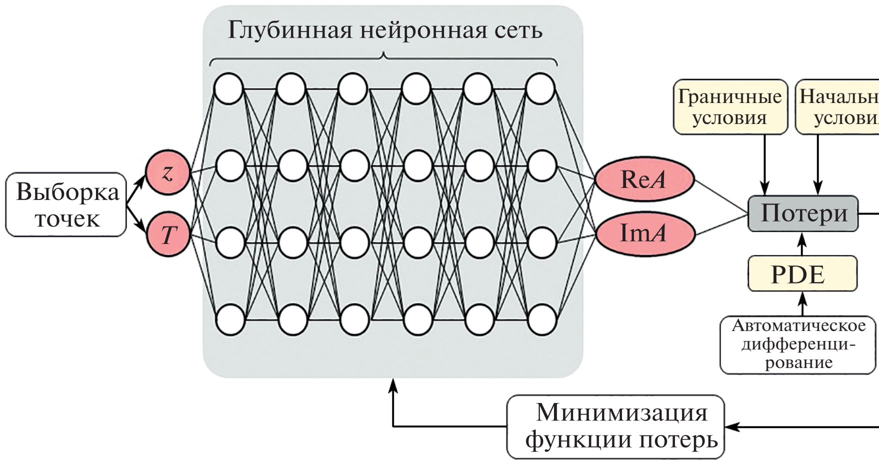

Рис. 1. Структура PINN для решения нелинейного уравнения Шредингера.

ся в диапазонах 1e-6 – 10 нДж и 5 – 10 000 м соответственно.

2.3. *Метод PINN*

Сформулируем метод PINN к нашей задаче. PINN является нейронной сетью, на вход которой подаются значения z и t. На выходе нейронной сети имеем действительную u и мнимую v части $A(z,t)$:

$$u, v = \mathrm{PINN}(z, T). \tag{15}$$

В процессе обучения минимизируется функция потерь, имеющая вид

$$E = E_{pde} + E_b + E_{in}, \tag{16}$$

где E_{pde} – невязка уравнения, рассчитанная в N_{pde} точках расчетной области,

$$E_{pde} = \sum_{j=1}^{N_{pde}} \left[\frac{\partial \mathrm{PINN}(z_j, T_j)}{\partial z} + \frac{i\beta_2}{2} \frac{\partial^2 \mathrm{PINN}(z_j, T_j)}{\partial T^2} + \frac{\alpha}{2} \mathrm{PINN}(z_j, T_j) - i\gamma \left| \mathrm{PINN}(z_j, T_j) \right|^2 \mathrm{PINN}(z_j, T_j) \right]^2, \tag{17}$$

а E_b и E_{in} – невязки граничных и начальных условий, соответственно, т.е.

$$E_b = \sum_{j=1}^{N_b} \left[\mathrm{PINN}(0, T_j) - A_0(T_j) \right]^2, \quad E_{in} = \sum_{j=1}^{N_{in}} \left[\mathrm{PINN}(z_j, \pm T_{\max}) \right]^2, \tag{18}$$

где N_b и N_{in} – количества точек, взятых на $z = 0$ и $T = \pm T_{\max}$ соответственно. Частные производные $\frac{\partial \mathrm{PINN}(z_j, T_j)}{\partial z}$ и $\frac{\partial^2 \mathrm{PINN}(z_j, T_j)}{\partial T^2}$ рассчитываются с помощью автоматического дифференцирования. Минимизация функции E производится по $N = N_{pde} + N_b + N_{in}$ точкам $\{z_j, t_j\}$, $j = 1, \ldots, N$ из расчетной области. Структура PINN показана на рис. 1.

Эффективность PINN зависит от количества слоев и нейронов в сети, функций активации, выбора точек, в которых производится расчет функции потерь. Кроме того, как отмечено в работе [15], точность метода можно повысить с помощью весов при слагаемых функции потерь:

$$E = \lambda_{pde} E_{pde} + \lambda_b E_b + \lambda_{in} E_{in}, \tag{19}$$

где λ_{pde}, λ_b и λ_{in} – соответствующие веса.

3. РЕЗУЛЬТАТЫ РАСЧЕТОВ

3.1. *Конфигурация PINN*

Мы разработали модель PINN на основе PyTorch. Нейронная сеть может включать линейные и ResNet-блоки с разными количеством нейронов и функциями активации. В работе [16] предлагается использовать оптимизатор Adam совместно с алгоритмом L-BFGS. Однако в наших расчетах мы обнаружили, что алгоритм Adamax имеет схожую эффективность по сравнению с комбинацией двух вышеупомянутых минимизаторов. Во всех рассмотренных сценариях функция активации Tanhshrink работает лучше, чем функции активации ReLU и Tanh. Поэтому, чтобы упростить анализ, во всех расчетах мы используем по умолчанию Adamax и Tanhshrink.

Фактическое значение скорости обучения минимизатора регулируется планировщиком. Он умножает текущее значение скорости обучения на коэффициент (по умолчанию 0.9), если функция потерь не уменьшается после определенного количества итераций (по умолчанию 30 итера-

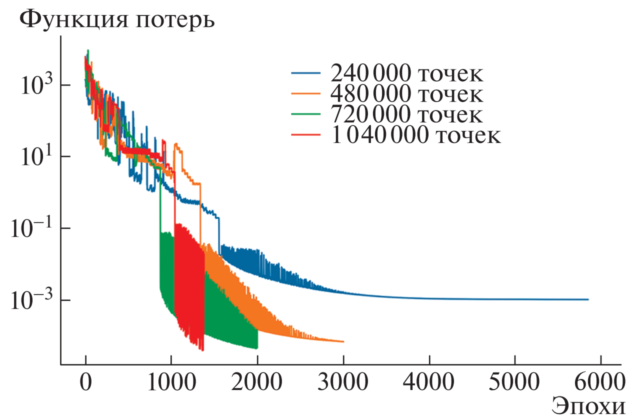

Рис. 2. Функция потерь при разном количестве точек сэмплирования.

ций). Минимальное значение скорости обучения установлено на уровне 1e-7. Частные производные вычисляются с помощью встроенных функций автоматического дифференцирования PyTorch. Веса сети инициализируются с помощью метода Ксавье. Обучение PINN производится на GPU NVidia Tesla V100 16GB. Если не указано иное, время обучения установлено на уровне двух часов для одного сценария. В модели могут быть использованы преобразования (8)–(11) и нормировка (12), а также балансировка слагаемых в функции потерь (19). Для оценки точности PINN используются референсные решения, полученные с помощью метода разделения шагов. Методом сеточного поиска было найдено, что для большинства сценариев наилучшие результаты PINN получаются при использовании конфигурации с 8 ResNet блоками по 150 нейронов в каждом блоке.

Количество точек сэмплирования существенно влияет на время обучения и точность модели. Время обучения одной эпохи линейно зависит от количества точек выборки. На рис. 2, иллюстрирующем функцию потерь, можно выделить области плавного снижения функции потерь, резкого снижения и области сильных колебаний с последующим плато. Увеличивая количество точек с 240000 до 720000, мы получаем улучшение результатов, а именно: потери уменьшаются, спад происходит раньше, обучение имеет тенденцию заканчиваться в стабильной точке до начала плато. Однако слишком большое количество точек (в нашем случае 1040000) приводит к негативным последствиям, а именно: спад наступает позже и обучение прекращается в области сильных колебаний функции потерь. Таким образом, при оптимальном количестве точек выборки процедура обучения останавливается, когда функция потерь достигает плато. В представленных расчетах используются 720000–960000 точек.

3.2. Применение PINN к нелинейному уравнению Шредингера

Применим описанную конфигурацию PINN к уравнению (1). Результаты показаны в табл. 1. Заметим, что с увеличением длины оптоволокна, а также с ростом энергии импульса теряется сходимость. С физической точки зрения это можно объяснить тем, что с ростом этих двух параметров решение для действительной и мнимой частей становится все более осциллирующим. А согласно частотному принципу глубокого обучения (F-principle) [17], наличие высоких частот в решении увеличивает время обучения. Недавно в работе [18] был обнаружен аналогичный эффект. Это наблюдение позволяет предположить, что существует максимальный уровень колебаний (приводящий к верхним ограничениям на длину волокна), выше которого PINN не сходится. В нашем случае увеличение числа слоев и количества нейронов не приводило к сходимости в проблемных случаях.

3.3. Влияние замены переменных на сходимость

Далее, применим PINN к уравнению (13), полученному с помощью преобразований (8)–(11) и нормализации (12). В этом случае сходимость получена для мощности 1 мВт при длине волокна 1 км (см. табл. 2 и рис. 3). Заметим, что точность решения PINN несколько ниже в тех областях,

Таблица 1. Результаты PINN для различных энергий и длин волокон в нелинейных случаях

Энергия/мощность	Длина волокна, м	Время обучения, мин	MSE	Loss
10 нДж/10 кВт	5	240	0.25	0.759
	15		Нет сходимости	
0.1 нДж/100 Вт	100	60	8.4e-4	1.4e-5
	200		Нет сходимости	
1e-3 нДж/1 Вт	100	30	3.5e-6	3.1e-7
	200		Нет сходимости	
1e-6 нДж/1 мВт	100		Нет сходимости	

Таблица 2. Ошибка решения PINN при решении уравнения (13)

Энергия/мощность	Длина волокна, м	Время обучения, мин	MSE	Loss
10 нДж/10 кВт	5	Нет сходимости		
	15	Нет сходимости		
0.1 нДж/100 Вт	100	60	0.00131	0.00021
	500	120	0.01070	0.00318
1e-3 нДж/1 Вт	100	30	6.126e-7	4.202e-5
	1000	60	1.76e-4	2.28e-3
1e-6 нДж/1 мВт	100	30	6.408e-10	1.976e-5
	1000	60	9.478e-8	1.26e-3

где наблюдаются сильные колебания, что опять же объясняется частотным принципом.

В табл. 3 сравнивается эффективность подхода PINN с эффективностью традиционных алгоритмов, а именно метода разделения шагов (split-step Foutier method – SSFM) и его распапаллелиной модификации (massive parallel algorithm – MPA). Детали реализации описаны в работе [19]. Расчеты выполнены с шагом интегрирования 5×10^{-4} м. Несмотря на то что цифры, приведенные в табл. 3, являются ориентировочными, можно сделать вывод, что время обучения PINN сопоставимо с временем вычислений традиционных методов, хотя и не превосходит результаты SSFM и MPA.

В линейном случае (в котором нелинейным слагаемым при коэффициенте γ пренебрегается) PINN сходится для всех энергий при длине оптического волокна до 2 км, как показано в табл. 4. При этом, несмотря на то, что функция потерь имеет относительно малые значения, ошибки в решении могут превышать 5%. В этом случае косвенным фактором, позволяющим оценить точность решения, является выполнение закона сохранения энергии импульса.

Сравнивая табл. 2 и 4, мы видим, что максимальные длины оптического волокна, при которых PINN еще сходится, различны при энергиях 10 и 0.1 нДж. Для понимания этого факта приведем численные значения L_n и L_d в табл. 5. Заметим, что если L_n меньше L_d, то именно L_n определяет максимальную длину волокна, которая зависит от энергии. Если $L_d < L_n$, то L_d определяет длину волокна, которая в этом случае не зависит от энергии. Эмпирическим путем мы получили, что область сходимости PINN по длине волокна определяется как $\sim\min(50L_d, 50L_n)$. Отметим, что во всех моделях параметр затухания $\alpha = 0$, что соответствует $L_a \to \infty$. При значениях $\alpha > 0$ сигнал затухает после прохождения расстояния $\sim L_a$ и тогда вопрос о сходимости PINN снимается.

Большая разница в коэффициентах при членах уравнения может вызвать численные трудности для PINN. Рассмотрим коэффициенты k_1, k_2, b и c, которые получаются после замены переменных (8)–(12). Численные значения приведены в табл. 6, из которой следует, что решение сходится, если коэффициенты при членах NLSE отличаются менее чем на 9 порядков. В противном случае вклад слагаемого при малом коэффициенте в функцию потерь при расчете E_{pde} становится пренебрежимо мал, и PINN не может корректно решить уравнение.

3.4. Синусоидальное отображение

Одной из возможных причин того, почему PINN не сходится в ряде случаев, является попадание в локальный минимум функции потерь. Для решения этой проблемы одним из перспективных подходов является использование синусоидального отображения входных данных [20], в котором синусоидальная функция активации,

$$\gamma(x) = \sin(2\pi(Wx + b)), \tag{20}$$

применяется на выходе первого слоя. В наших расчетах мы видели, что для решения нелинейного уравнения Шредингера более точные результаты получаются, если разделить первый слой на узлы, относящиеся к x и t, и применить (20) только к узлам t. Соответствующая структура PINN показана на рис. 4. При этом значения MSE уменьшаются в среднем в два раза. Здесь интересно провести аналогию с методом SSFM, в котором применяется преобразование Фурье к координате t. Ранее в работе [21] было предложено ис-

Таблица 3. Времена счета методов SSFM, MPA и PINN для длины оптоволокна 1 км

Алгоритм	SSFM	MPA	PINN
Платформа	C++/CPU	C++/GPU	PyTorch/GPU
Время счета, часы	0.8	0.4	1

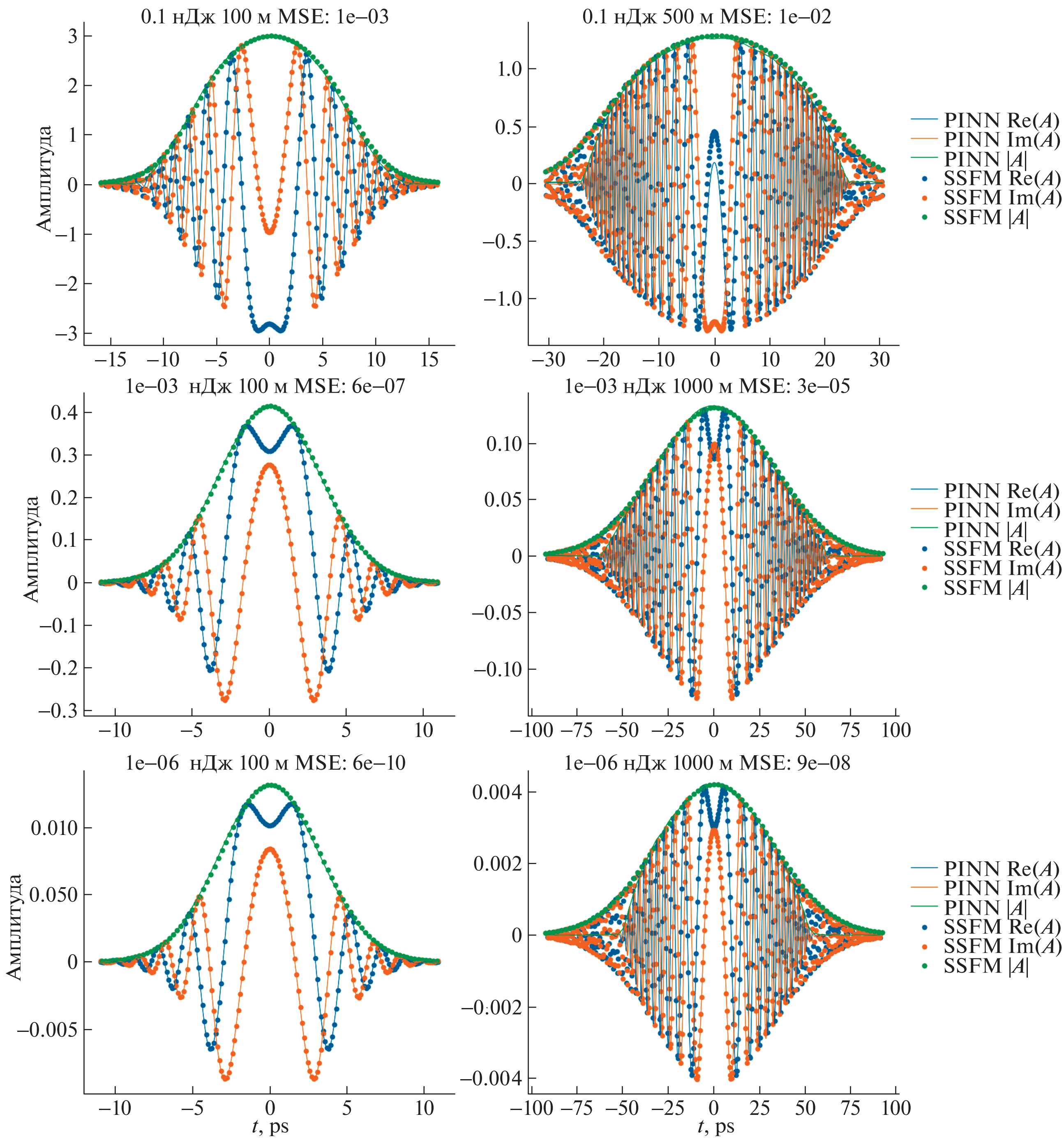

Рис. 3. Сравнение результатов PINN с применением нормировки с результатами традиционного SSFM для энергий 0.1, 1e-03, 1e-06 нДж при длинах волокна 100, 500, 1000 м.

пользовать блоки преобразования Фурье для повышения точности PINN. Таким образом, использование синусоидальной функции активации для t может быть частично оправдано, а при проектировании архитектуры нейронной сети можно руководствоваться математическими свойствами задачи, как предложено в работе [22].

3.5. Подбор весов к слагаемым функции потерь

Рассмотрим влияние коэффициентов при слагаемых функции потерь. Сразу заметим, что адаптивный алгоритм подборки весов, описанный в работе [15], не привел к существенному улучшению результатов. Поэтому мы рассмотрим под-

Таблица 4. Результаты PINN с применением нормировки для различных энергий и длин волокон в линейных случаях

Энергия/мощность	Длина волокна, км	MSE	Loss	Сохранение энергии
10 нДж/10 кВт	1	2.0759	0.0019	98%
	2	1.096	0.0031	93%
	10	Нет сходимости		13%
0.1 нДж/100 Вт	1	0.0125	0.0011	98%
	2	0.0109	0.0031	93%
	10	Нет сходимости		13%
1e-3 нДж/1 Вт	1	2e-4	0.0013	98%
	2	1.096e-4	0.0031	93%
	10	Нет сходимости		13%
1e-6 нДж/1 мВт	1	1.396e-7	0.0012	98%
	2	1.091e-7	0.0031	93%
	10	Нет сходимости		13%

Таблица 5. Максимальные длины оптоволокна и соответствующие длины нелинейности (дисперсионная длина во всех случаях равна L_d = 18.8 м)

Нелинейность	Энергия/мощность	Макс. длина волокна, м	L_n, м	Ограничение по длине
Да	10 нДж/10 кВт	~ 5	0.09284	$50L_n$
	0.1 нДж/100 Вт	~ 500	9.28475	$50L_n$
	1e-3 нДж/1 Вт	~ 1000	928.475	$50L_d$
	1e-6 нДж/1 мВт	~ 1000	928475	$50L_d$
Нет	10 нДж/10 кВт	~ 1000	–	$50L_d$
	0.1 нДж/100 Вт	~ 1000	–	$50L_d$
	1e-3 нДж/1 Вт	~ 1000	–	$50L_d$
	1e-6 нДж/1 мВт	~ 1000	–	$50L_d$

бор веса λ перед невязкой по начальным условиям:

$$E = E_{pde} + E_b + \lambda E_{in}. \tag{21}$$

В процессе обучения мы меняем λ так, чтобы E_{pde} и λE_{in} были одного порядка, что позволяет не только уменьшить невязку по начальным условиям, но и в целом повысить точность. На рис. 5 показан результат совместной работы синусоидального отображения и подбора λ. Благодаря их использованию, PINN может воспроизводить области с высокочастотными колебаниями на границах окна, уменьшая на порядок средне-квадратичную ошибку решения.

3.6. *Meta-PINN*

Все вышеперечисленные PINN были обучены для одного сценария. Это означает, что при изменении любого параметра, например, энергии импульса, PINN необходимо переобучить. Этот процесс может занимать до двух часов на GPU для одного сценария. При такой скорости PINN не может превзойти SSFM, который занимает менее 1 ч на CPU (см. табл. 3). Однако PINN способен решать параметрические уравнения. Такой тип PINN часто называют мета-PINN [23]. В этом случае мета-PINN имеет дополнительные входы для параметров уравнения.

Рассмотрим задачу, в которой энергия входного импульса может изменяться в диапазоне от 1e-3 до 10e-3 нДж. Результаты для нескольких энергий приведены на рис. 6. С увеличением энергии импульса нелинейные эффекты становятся более выраженными, но это не мешает PINN предсказывать решение с приемлемой точностью. Отметим, что мета-PINN обладает хорошей обобщающей способностью. Хотя обучение проводилось только для энергий импульса $k \times 10^{-3}$ нДж, где k – целое число, мета-PINN дает точные результаты для

Таблица 6. Коэффициенты нормировки

Энергия/мощность	Длина волокна, м	Врем. окно, пс	k_1	k_2	b	c	Решение
10 нДж/10 кВт	5	100	0.3	**166.5**	**1.8e-7**	53.8	**Нет**
	15	100	0.8	**166.5**	**5.5e-7**	161.6	**Нет**
	100	100	5.3	166.5	3.7e-6	–	Да
	300	100	15.9	166.5	1.1e-5	–	Да
	1000	300	53.1	499.5	4.1e-6	–	Да
	2000	600	106.3	999	2.0e-6	–	Да
	10000	2000	531.3	**3330**	**9.2e-7**	–	**Нет**
0.1 нДж/100 Вт	100	100	5.3	166.5	3.7e-6	10.8	Да
	500	200	26.6	333.0	4.6e-6	53.9	Да
	1000	300	53.1	499.5	4.1e-6	–	Да
	2000	600	106.3	999	2.0e-6	–	Да
	10000	2000	531.3	**3330**	**9.2e-7**	–	**Нет**
1e-3 нДж/1 Вт	100	100	5.3	166.5	3.7e-6	0.1	Да
	1000	300	53.1	499.5	4.1e-6	–	Да
	2000	600	106.3	999	2.0e-6	–	Да
	10000	2000	531.3	**3330**	**9.2e-7**	–	**Нет**
1e-6 нДж/1 мВт	100	100	5.3	166.5	3.7e-6	1.1e-4	Да
	1000	300	53.1	499.5	4.1e-6	–	Да
	2000	600	106.3	999	2.0e-6	–	Да
	10000	2000	531.3	**3330**	**9.2e-7**	–	**Нет**

**Жирным шрифтом выделены случаи, в которых разница в коэффициентах может быть причиной отсутствия сходимости*

промежуточных значений энергии. Процедура обучения мета-PINN занимает примерно в 2–3 раза больше времени, чем PINN для одного сценария, в то время как предсказание одного сценария с помощью мета-PINN занимает лишь доли секунды. Это говорит о том, что мета-PINN может быть более эффективным, чем традиционные методы (в частности, метод разделения шагов),

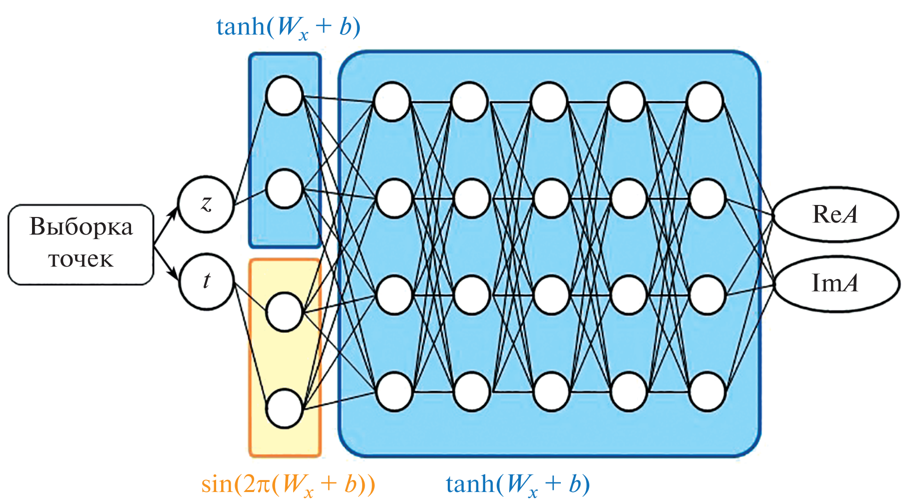

Рис. 4. Архитектура PINN при частичном использовании синусоидальной функции активации.

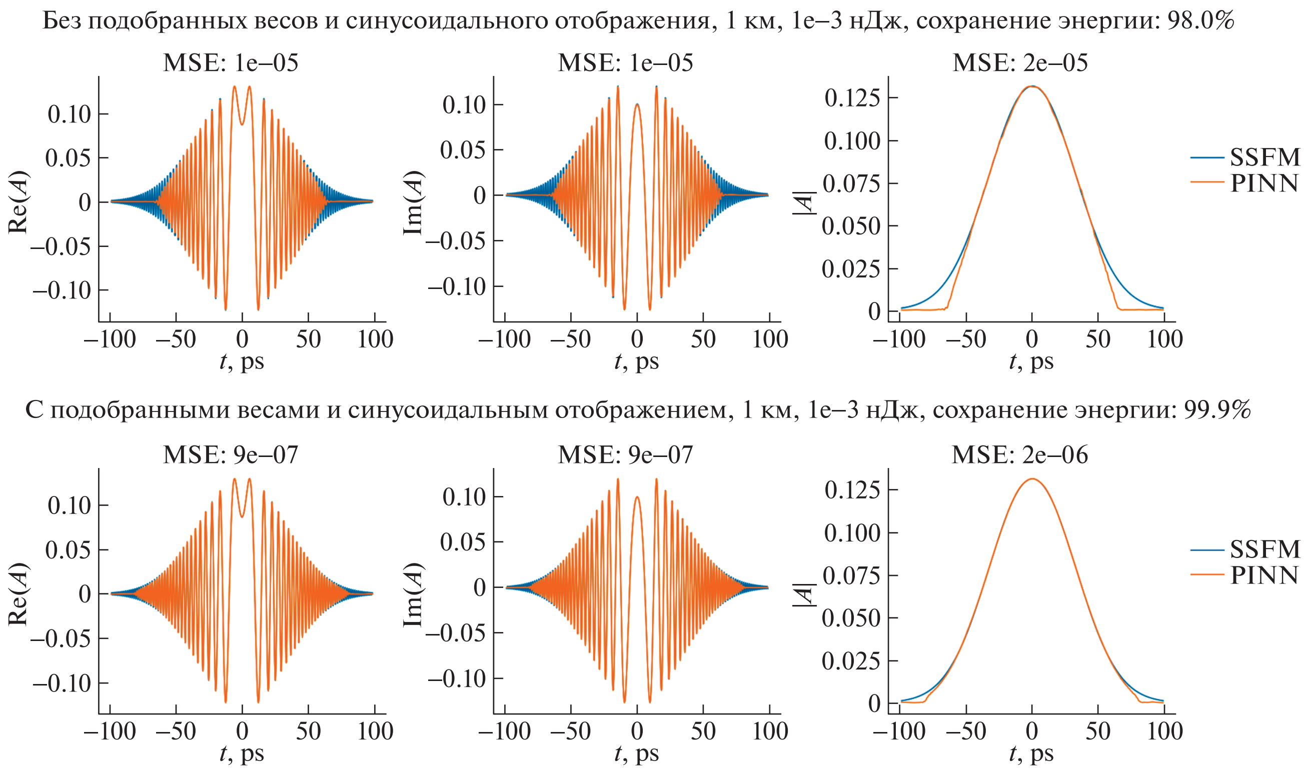

Рис. 5. Улучшение результатов PINN после применения подобранных весов и синусоидального отображения входных данных.

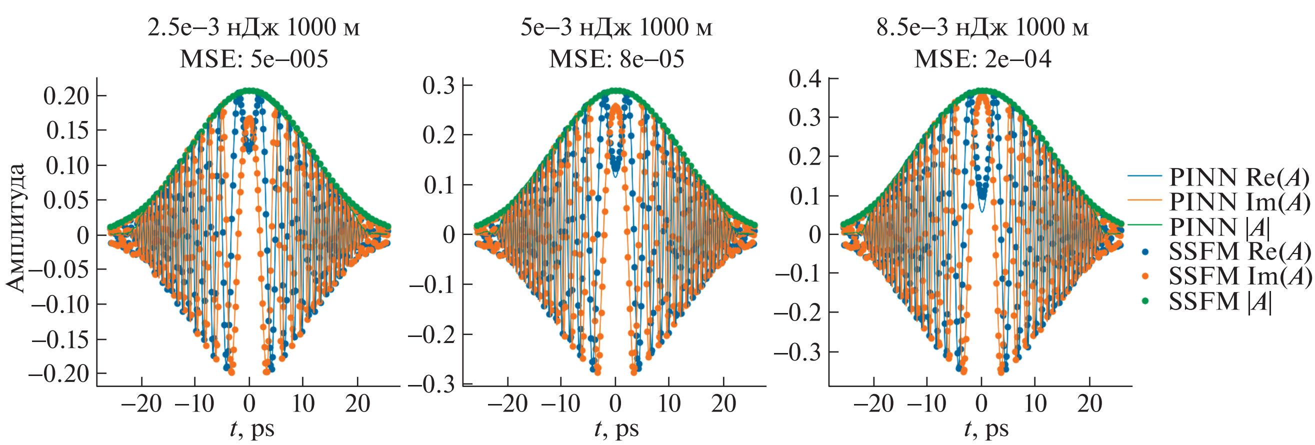

Рис. 6. Сравнение результатов мета-PINN с результатами традиционного SSFM для следующих энергий: 2.5e-3, 5e-3, 8.5e-3 нДж.

которые могут решать только один сценарий за раз.

4. ЗАКЛЮЧЕНИЕ

В данной работе было показано, что коэффициенты перед членами уравнения Шредингера определяют эффективность решения с помощью метода PINN. С помощью замены переменных можно улучшить сходимость PINN. Однако при мощности импульса выше 10 кВт такое преобразование ухудшает сходимость. PINN устойчиво сходится, если коэффициенты перед слагаемыми уравнения отличаются менее чем на 9 порядков. Замена переменных несколько выравнивает коэффициенты, тем самым улучшая сходимость PINN.

Скорость сходимости PINN уменьшается из-за колебаний в действительной и мнимой частях решения. Максимальная длина волокна для PINN ограничена характерными длинами, такими как дисперсионная длина L_d и нелинейная длина L_n. Эмпирически установлено, что устойчивые результаты могут быть получены при длинах распространения, меньших $\min(50L_d, 50L_n)$. При длинах распространения, превышающих это значение, решение имеет сильные осцилляции, что создает трудности для PINN в соответствии с F-принципом. Было показано, что малые значения функции потерь не являются надежным индикатором точности PINN. В тех случаях, когда длина волокна близка к $\min(50L_d, 50L_n)$, точность решения можно оценить по сохранению энергии.

Балансировка членов функции потерь может повысить точность. В процессе обучения весовые коэффициенты должны быть адаптированы таким образом, чтобы привести слагаемые к одному порядку величины.

Использование синусоидальной функции активации в первом слое также улучшает результат. При увеличении длины волокна имеет смысл разделить первый слой на два подслоя и применять синусоидальную функцию активации только для t-координаты. Совместное использование этих методов позволяет справиться с высокочастотными колебаниями, усиливающимися к краям временного окна, и улучшить сохранение энергии до 99.9%.

Была успешно применена концепция мета-PINN к задаче с произвольным значением мощности импульса. Хотя PINN в наших расчетах оказался медленнее метода разделения шагов, концепция мета-PINN позволяет предобучить модель и использовать ее для разных сценариев, что делает такой метод конкурентоспособным.

СПИСОК ЛИТЕРАТУРЫ

1. *Kollmannsberger S., D'Angella D., Jokeit M., Herrmann L.* "Physics-Informed Neural Networks". В: Deep Learning in Computational Mechanics. Springer International Publishing, 2021. P. 55–84. https://doi.org/10.1007/978-3-030-76587-3_5
2. *Markidis S.* "The Old and the New: Can Physics-Informed Deep-Learning Replace Traditional Linear Solvers?" В: Frontiers in Big Data 4. 2021. issn: 2624-909X. https://doi.org/10.3389/fdata.2021.669097
3. *Govind P Agrawal*. Fiber-Optic Communication Systems. Wiley, 2021. https://doi.org/10.1002/9781119737391
4. *Monterola C., Saloma C.* "Solving the nonlinear Schrodinger equation with an unsupervised neural network". en. В: Opt. Express 9.2. 2001. P. 72–84.
5. *Raissi M., Perdikaris P., Karniadakis G.E.* "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations". В: Journal of Computational Physics. 2019. V. 378. P. 686–707.
6. *Chuprov I., Efremenko D., Gao J., Anisimov P., Zemlyakov V.* Scaling transformation of the multimode nonlinear Schrödinger equation for physics-informed neural networks. 2022. arXiv: 2209.14641 [cs.NE].
7. *Karniadakis G.E., Kevrekidis I.G., Lu Lu, Perdikaris P., Wang S., Liu Yang*. "Physics-informed machine learning". en. В: Nat Rev Phys 3.6. 2021. P. 422–440.
8. *Xiaotian Jiang, Danshi Wang, Qirui Fan, Min Zhang, Chao Lu*, Alan Pak Tao Lau. "Solving the Nonlinear Schrödinger Equation in Optical Fibers Using Physics-informed Neural Network". В: 2021 Optical Fiber Communications Conference and Exhibition (OFC). 2021. P. 1–3.
9. *Jagtap A.D., Karniadakis G.E.* "Extended Physics-Informed Neural Networks (XPINNs): A Generalized Space-Time Domain Decomposition Based Deep Learning Framework for Nonlinear Partial Differential Equations". В: Communications in Computational Physics 28.5. 2020. P. 2002–2041. https://doi.org/10.4208/cicp.oa-2020-0164
10. *Lee J.D., Simchowitz M., Jordan M.I., Recht B.* Gradient Descent Converges to Minimizers. 2016. eprint: arXiv:1602.04915.
11. *Shamsulhaq Basir*, *Inanc Senocak*. "Critical Investigation of Failure Modes in Physics-informed Neural Networks". В: AIAA SCITECH 2022 Forum. American Institute of Aeronautics и Astronautics, 2022. https://doi.org/10.2514/6.2022-2353
12. *Yeonjong Shin*. "On the Convergence of Physics Informed Neural Networks for Linear Second-Order Elliptic and Parabolic Type PDEs". В: Communications in Computational Physics 28.5. 2020. P. 2042–2074. https://doi.org/10.4208/cicp.oa-2020-0193
13. *Grossmann T.G., Komorowska U.J., Latz J.*, Carola-Bibiane Schönlieb. Can Physics-Informed Neural Networks beat the Finite Element Method? 2023. arXiv: 2302.04107 [math.NA].
14. *Yubin Zang, Zhenming Yu, Kun Xu, Minghua Chen, Sigang Yang, Hongwei Chen*. "Universal Fiber Models based on PINN Neural Network". В: Asia Communications and Photonics Conference 2020. M4A.266. OSA. 2020.
15. *Sifan Wang, Xinling Yu и Paris Perdikaris*. "When and why PINNs fail to train: A neural tangent kernel perspective". В: Journal of Computational Physics 449. 2022. P. 110768. https://doi.org/10.1016/j.jcp.2021.110768
16. *Xiaotian Jiang, Danshi Wang, Xue Chen, Min Zhang.* "Physics-Informed Neural Network for Optical Fiber Parameter Estimation from the Nonlinear Schrödinger Equation". В: Journal of Lightwave Technology. 2022. P. 1–11. https://doi.org/10.1109/jlt.2022.3199782
17. *Tao Luo*. "Theory of the Frequency Principle for General Deep Neural Networks". В: CSIAM Transactions on Applied Mathematics 2.3. 2021. P. 484–507. https://doi.org/10.4208/csiam-am.so-2020-0005
18. *Nasim Rahaman, Aristide Baratin, Devansh Arpit, Felix Draxler, Min Lin, Fred A. Hamprecht, Yoshua Bengio,*

Aaron Courville. On the Spectral Bias of Neural Networks. 2019. arXiv: 1806.08734 [stat.ML].

19. *Kazakov E., Jiexing Gao, Anisimov P., Zemlyakov V.* "Parallelization of the Generalized Multimode Nonlinear Schrödinger Equation Solver: A Performance Analysis". B: Communications in Computer and Information Science. Springer Nature Switzerland, 2023. P. 137–151. https://doi.org/10.1007/978-3-031-38864-4_10
20. *Jian Cheng Wong, Chinchun Ooi, Abhishek Gupta, Yew-Soon Ong*. "Learning in Sinusoidal Spaces with Physics-Informed Neural Networks". B: IEEE Transactions on Artificial Intelligence. 2022. P. 1–15. https://doi.org/10.1109/tai.2022.3192362
21. *Sifan Wang, Hanwen Wang*, Paris Perdikaris. "On the eigenvector bias of Fourier feature networks: From regression to solving multi-scale PDEs with physics-informed neural networks". B: Computer Methods in Applied Mechanics and Engineering 384. 2021. P. 113938. https://doi.org/10.1016/j.cma.2021.113938
22. *Hager Ch., Pfister H.D.* "Deep Learning of the Nonlinear Schrödinger Equation in Fiber-Optic Communications". B: 2018 IEEE International Symposium on Information Theory (ISIT). IEEE, июнь 2018. https://doi.org/10.1109/isit.2018.8437734.
23. *Xiaotian Jiang, Danshi Wang, Xue Chen, Min Zhang*. "Physics-Informed Neural Network for Optical Fiber Parameter Estimation from the Nonlinear Schrödinger Equation". B: Journal of Lightwave Technology. 2022. P. 1–11. https://doi.org/10.1109/jlt.2022.3199782
24. *Psaros A.F., Kawaguchi K., Karniadakis G.E.* "Meta-learning PINN loss functions". B: CoRR abs/2107.05544 (2021). arXiv: 2107.05544.

OPTIMIZATION OF PHYSICS-INFORMED NEURAL NETWORKS FOR NONLINEAR SCHRÖDINGER EQUATION

I. A. Chuprov[a], J. Gao[a], D. S. Efremenko[a], E. A. Kazakov[a], F. A. Buzaev[a], and V. V. Zemlyakov[a]

[a] *Moscow Research Center, 2012 Labs, Huawei Technologies Co., Ltd., Moscow, Russia*

Presented by Academician of the RAS A.L. Semenov

In this paper, PINN is applied to the NLSE equation in order to determine the performance range and limiting factors. Some tools, such as manual weights of the loss function components, and selective application of the sinusoidal activation function, are applied to improve the results. Accepting the fact that PINN loses to SSFM in terms of performance, the application of Meta-PINN to NLSE is investigated to cover the range of parameters, demonstrating the successful generalisation ability of Meta- PINN. The paper concludes with a recommendation on how to tune PINN to successfully solve NLSE.

Keywords: physics-informed neural networks, Nonlinear Schrödinger equation, nonlinear fiber optics, fine-tuning neural networks