Московский государственный университет имени М.В. Ломоносова
Факультет вычислительной математики и кибернетики
Кафедра математических методов прогнозирования

Юдин Никита Евгеньевич

Вариационный вывод в нейронных стохастических дифференциальных уравнениях

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

Научный руководитель:
д.ф-м.н., профессор, академик РАН
Рудаков Константин Владимирович

Научный консультант:
д.ф-м.н., доцент, профессор РАН
Воронцов Константин Вячеславович

Москва, 2020

Содержание

1 Введение

Байесовский вывод является важным инструментом современной статистики, однако огромное количество важных и интересных задач с практической стороны препятствуют прямому применению точного байесовского вывода. И для решения подобных задач часто приходится проводить различного рода релаксации с помощью методов приближённого байесовского вывода, в частности, вариационного [1]. Процедура вариационного вывода благодаря своей масштабируемости и эффективности заслужила популярность в современных исследованиях вероятностных моделей, использующих практически полезные аппроксимационные возможности нейронных сетей. Данные модели могут задаваться как в явном виде с аналитически или численно моделируемыми функциями правдоподобия [2, 3], так и в неявном виде, требующим при построении только возможность сэмплирования [4, 5], при этом, также активно развивается направление, соединяющее в себе оба подхода к моделированию [6, 7].

Однако важным аспектом вариационного вывода является то, насколько близко может быть настраиваемое вариационное приближение точного апостериорного распределения в семействе выбранных моделей. Во многом качество итогового решения зависит от начального приближения и богатства выбранного семейства вариационных приближений. Добиться расширения семейства вариационных приближений можно прямым усложнением содержащихся в рассматриваемой модели отображений, например, увеличением количества слоёв в нейронной сети легче повысить экспрессивность модели, чем увеличением размеров каждого слоя [8]. Также повысить качество решения задачи можно с помощью частичной модификации или полного изменения спецификации рассматриваемой модели [9].

На практике часто при выборе более естественной модели для исследуемого объекта или процесса происходит заметный рост качества решения в рассматриваемых задачах. Для различного рода динамических систем такими естественными моделями нередко выступают те, что частично или полностью описываются системами обыкновенных (ОДУ) или стохастических (СДУ) дифференциальных уравнений [10]. В свою очередь СДУ, будучи обобщением ОДУ, являются естественным инструментом описания явлений, вызванных огромным количеством малых ненаблюдаемых невооружённым глазом взаимодействий, таких, как динамика жидкости и газа [11], изменение цен на финансовых рынках [12]. Однако применение СДУ в математическом моделировании представляет собой определённые проблемы, связанные с плохой масштабируемостью многих существующих решений как по времени работы при моделировании, так и по требуемой памяти [13]. Также определённые трудности возникают при попытке адекватно оценить диффузию без редукции СДУ к ОДУ [14]. Данная работа посвящена развитию методов настройки вероятностных моделей, параметризованных стохастическими дифференциальными уравнениями, в том числе и в рамках процедуры вариационного вывода. В работе представлен обзор использования в различных задачах машинного обучения моделей, параметризованных ОДУ и СДУ. Далее обоснованно приводится проце-

дура оценки производных по параметрам в моделях с дифференциальными уравнениями в рамках градиентных методов оптимизации. После выводится обобщение процедуры оценки параметров СДУ в вариационном выводе.

2 Постановка задачи

В работе решается задача вероятностного моделирования:

- дана выборка из N_D независимо распределённых последовательностей

$$X^{N_D} = \{({}^j x_{t_1},\ \ldots,\ {}^j x_{t_{N_j}})\}_{j=1}^{N_D},\ ({}^j x_{t_i}, t_i) \in \mathbb{X} \times [0,T],\ t_i < t_{i+1};$$

  $\mathbb{X}$ — пространство элементов последовательности, каждая последовательность является реализацией случайного процесса во времени $[0,T]$, то есть временным рядом;
- для каждого элемента последовательности ${}^j x_{t_i} \in \mathbb{X}$ задано описание в виде вектора–столбца в признаках $f_j,\ j = \overline{1,d_x}$:

  ${}^j\mathbf{x}_{t_i} = (f_1({}^j x_{t_i}),\ \ldots,\ f_{d_x}({}^j x_{t_i}))^{\mathrm{T}} \in \mathbb{D} = \mathbb{D}_{f_1} \times \cdots \times \mathbb{D}_{f_{d_x}}$, $\mathbb{D}$ — признаковое пространство;

- введено обозначение $\mathrm{p}^* :\ \mathbb{D} \times \cdots \times \mathbb{D} \to [0, +\infty)$ — неизвестная плотность вероятности;
- необходимо построить приближение p истинной p^* по выборке.

Критерии качества:

- средний логарифм правдоподобия:

$$\log \mathrm{p}(\hat{X}^{N_D}) = \frac{1}{N_D} \sum_{j=1}^{N_D} \log \mathrm{p}(({}^j\hat{\mathbf{x}}_{t_1},\ \ldots,\ {}^j\hat{\mathbf{x}}_{t_{N_j}})),$$

  $\mathrm{p}(\cdot)$ — плотность вероятности;
- квадратный корень среднеквадратической ошибки модели:

$$rmse(X^{N_D}, \hat{X}^{N_D,K}) = \sqrt{\frac{1}{N_D K} \sum_{j,k=1}^{N_D,K} \frac{1}{N_j} \sum_{i=1}^{N_j} \frac{1}{d_x} \left\| {}^j\mathbf{x}_{t_i} - {}^j\hat{\mathbf{x}}^k_{t_i} \right\|_2^2},$$

  $X^{N_D} = \{({}^j x_{t_1},\ \ldots,\ {}^j x_{t_{N_j}})\}_{j=1}^{N_D}$ — целевая выборка, $\hat{X}^{N_D,K} = \{({}^j\hat{x}^k_{t_1},\ \ldots,\ {}^j\hat{x}^k_{t_{N_j}})\}_{j,k=1}^{N_D,K}$ — сэмплы вероятностной модели p.

Решение задачи мотивировано следующими целями исследования в данной работе:

1) Разработать схемы настройки нейронных сетей, параметризованных с помощью стохастических дифференциальных уравнений (раздел 4).

2) Разработать и исследовать схемы вариационного вывода с α–Реньи дивергенциями для нейронных сетей, параметризованных с помощью стохастических дифференциальных уравнений (раздел 5).

3 Обзор и основные понятия

3.1 Нейронные обыкновенные дифференциальные уравнения

В нескольких работах исследователи заметили, что такие нейросетевые архитектуры, как *нейронные сети с пробрасыванием связи (residual networks)*, рекуррентные нейросетевые декодировщики, нормализующие потоки рекуррентно осуществляют преобразование скрытого состояния:

$$\mathbf{z}_{i+1} = \mathbf{z}_i + f(\mathbf{z}_i, \theta_i),$$

где $i \in \{0, \ldots, N_L - 1\}$ и $(\mathbf{z}_i, \theta_i) \in \mathbb{R}^{d_\mathbf{z}} \times \mathbb{R}^{d_{\theta_i}}$, (f, θ_i) — отображение для скрытого состояния $\mathbf{z}_i$ на слое i, N_L — количество слоёв. [3, 15, 16]. Данные рекуррентные процедуры можно рассмотреть как реализацию разностной схемы для решения обыкновенного дифференциального уравнения (ОДУ) методом Эйлера [17–19].

Таким образом, при уменьшении шага метода Эйлера (равносильно увеличению слоёв в нейросетевой модели) и при выполнении условий липшицевости функции f нейронную сеть можно задать с помощью следующей задачи Коши с гарантией существования и единственности решения:

$$\begin{cases} \dfrac{\mathrm{d}\,\mathbf{z}(t)}{\mathrm{d}\,t} = f(\mathbf{z}(t), t, \theta), \ t \in [0, T], \ \theta \in \mathbb{R}^{d_\theta}; \\ \mathbf{z}(0) = \mathbf{z}_0 \in \mathbb{R}^{d_\mathbf{z}} \text{ — вход нейронной сети.} \end{cases} \tag{1}$$

$f : \mathbb{R}^{d_\mathbf{z}} \times \mathbb{R}^+ \times \mathbb{R}^\theta \to \mathbb{R}^{d_\mathbf{z}}$, функция f непрерывна по Липшицу, $\mathbf{z}(T)$ — выход нейронной сети. В статье [20] обозначено несколько достоинств данного подхода.

- **Эффективность по памяти**. Задание нейронной сети в виде задачи Коши позволяет сократить эффективную занимаемую память моделью по сравнению с дискретным аналогом, так как возможно алгоритм распространения ошибки реализовать в виде динамики, заданной с помощью другой задачи Коши. То есть сеть занимает O(1) памяти как функция количества узлов в разностной схеме решения уравнения.

- **Адаптивность вычислений**. Заключается в свободе выбора численного метода решения дифференциального уравнения и в частоте дискретизации разностной схемы.

- **Эффективность параметризации**. При дискретизации решения задачи Коши параметры в соседних узлах не сильно отличаются по значению, что, как заявляют авторы

в [20], сокращает необходимое количество параметров в f для достижения требуемого качества.

- **Масштабируемые и обратимые нормализующие потоки** [21]. Данные модели, параметризованные дифференциальным уравнением, легче обучать, чем дискретные аналоги, также их можно обучать напрямую методом максимизации правдоподобия.

- **Модели непрерывных временных рядов** [10]. Предложенная параметризация модели позволяет более естественным образом обрабатывать данные с произвольным временным лагом.

3.1.1 Настройка параметров модели

Настройка параметров в модели (1) реализована с помощью решения вспомогательного дифференциального уравнения, сформулированного в *принципе максимума Понтрягина* [22]. Для вывода соответствующих формул рассмотрим дважды дифференцируемый липшицевый оптимизируемый критерий (функция потерь) $L(\cdot)$ с дважды дифференцируемой липшицевой функцией f:

$$L(\mathbf{z}(T)) = L\left(\mathbf{z}(0) + \int\limits_0^T f(\mathbf{z}(t), t, \theta)\,\mathrm{d}\,t\right).$$

В работе используется следующая нотация при записи производных и дифференциалов:

$$\mathrm{d}\,\mathbf{z}(t) = \left((\mathrm{d}\,\mathbf{z}_l(t))_{l=1}^{d_\mathbf{z}}\right)^\mathrm{T} \text{ — вектор–столбец;}$$

$$\frac{\mathrm{d}\,\mathbf{z}(t)}{\mathrm{d}\,t} = \left(\left(\frac{\mathrm{d}\,\mathbf{z}_l(t)}{\mathrm{d}\,t}\right)_{l=1}^{d_\mathbf{z}}\right)^\mathrm{T} \in \mathbb{R}^{d_\mathbf{z}\times 1} \cong \mathbb{R}^{d_\mathbf{z}} \text{ — вектор–столбец;}$$

$$\frac{\mathrm{d}\,L(\mathbf{z}(T))}{\mathrm{d}\,\mathbf{z}(t)} = \left(\frac{\mathrm{d}\,L(\mathbf{z}(T))}{\mathrm{d}\,\mathbf{z}_l(t)}\right)_{l=1}^{d_\mathbf{z}} \in \mathbb{R}^{1\times d_\mathbf{z}} \cong \mathbb{R}^{d_\mathbf{z}} \text{ — вектор–строка;}$$

$$\frac{\mathrm{d}\,L(\mathbf{z}(T))}{\mathrm{d}\,\theta} = \left(\frac{\mathrm{d}\,L(\mathbf{z}(T))}{\mathrm{d}\,\theta_l}\right)_{l=1}^{d_\theta} \in \mathbb{R}^{1\times d_\theta} \cong \mathbb{R}^{d_\theta} \text{ — вектор–строка;}$$

$$\frac{\partial f(\mathbf{z}(t), t, \theta)}{\partial \mathbf{z}(t)} = \left(\frac{\partial f_p(\mathbf{z}(t), t, \theta)}{\partial \mathbf{z}_q(t)}\right)_{p,q=1}^{d_\mathbf{z}} \in \mathbb{R}^{d_\mathbf{z}\times d_\mathbf{z}} \text{ — квадратная матрица;}$$

$$\frac{\partial f(\mathbf{z}(t), t, \theta)}{\partial \theta} = \left(\frac{\partial f_p(\mathbf{z}(t), t, \theta)}{\partial \theta_q}\right)_{p,q=1}^{d_\mathbf{z}, d_\theta} \in \mathbb{R}^{d_\mathbf{z}\times d_\theta} \text{ — прямоугольная матрица.}$$

Такая нотация введена из соображений удобства, единственный принципиальный нюанс, связанный с различием векторов–строк и векторов–столбцов, в данной работе состоит в том, что при оптимизации параметров θ градиентными методами обновление θ производится вдоль транспонированного анти(суб)градиента функции потерь. Производные функции потерь нейронной сети по $\mathbf{z}_0$ и по θ выводятся из решения вспомогательной задачи Коши:

$$\begin{cases} \dfrac{\mathrm{d}\,\mathbf{a}(t)}{\mathrm{d}\,t} = -\mathbf{a}(t)\dfrac{\partial f(\mathbf{z}(t), t, \theta)}{\partial \mathbf{z}(t)}, \ \dfrac{\mathrm{d}\,\mathbf{a}(t)}{\mathrm{d}\,t} \text{ — вектор–строка;} \\ \mathbf{a}(T) = \dfrac{\mathrm{d}\,L(\mathbf{z}(T))}{\mathrm{d}\,\mathbf{z}(T)}, \ \mathbf{a}(t) \text{ — вектор–строка.} \end{cases}$$

Производная по входу в нейронную сеть выражается следующим образом [20]:

$$\begin{cases} \mathbf{a}(t) = \dfrac{\mathrm{d}\,L(\mathbf{z}(T))}{\mathrm{d}\,\mathbf{z}(t)}, \ \mathbf{a}(t) \text{ — вспомогательная переменная для } \mathbf{z}(t); \\ \dfrac{\mathrm{d}\,L(\mathbf{z}(T))}{\mathrm{d}\,\mathbf{z}_0} = \mathbf{a}(0) = \mathbf{a}(T) + \int\limits_0^T \mathbf{a}(t)\dfrac{\partial f(\mathbf{z}(t), t, \theta)}{\partial \mathbf{z}(t)}\,\mathrm{d}\,t. \end{cases}$$

Производная по параметрам нейронной сети вычисляется по формуле [20]:

$$\frac{\mathrm{d}\,L(\mathbf{z}(T))}{\mathrm{d}\,\theta} = \int\limits_0^T \mathbf{a}(t)\frac{\partial f(\mathbf{z}(t), t, \theta)}{\partial \theta}\,\mathrm{d}\,t.$$

Данные уравнения позволяют нейросетевую модель (1) оптимизировать (суб)градиентными методами.

3.1.2 Временные ряды

Вероятностное моделирование временных рядов в нейронных ОДУ можно организовать следующим образом, предложенным в [20]. Допустим, необходимо для моментов времени $(t_1, \ \ldots, \ t_N)$ получить значения случайных величин из моделируемого временного ряда $(\mathbf{x}_{t_i})_{i=1}^N$. Первая часть данных в хронологическом порядке соответствует известным наблюдениям $(\mathbf{x}_{t_i})_{i=1}^n, \ n < N$, оставшиеся наблюдения $(\mathbf{x}_{t_{n+1}}, \ \ldots, \ \mathbf{x}_{t_N})$ необходимо промоделировать. Данную подзадачу вероятностного моделирования будем называть *экстраполяцией временного ряда*. В работе также рассматривается другой вид подзадачи вероятностного моделирования — *интерполяция временного ряда*. Она заключается в построении вероятностной модели для последовательности $(\mathbf{x}_{t_i})_{i=1}^N$, с помощью которой возможно описать динамику $\mathbf{x}_t$ в произвольный момент времени $t \in [0, T]$. Нейронное ОДУ можно ввести в качестве описания динамики латентной переменной $\mathbf{z}_{t_i} \in \mathbb{R}^{d_\mathbf{z}}$, соответствующей $\mathbf{x}_{t_i} \in \mathbb{R}^{d_\mathbf{x}}$:

$$\begin{cases} \mathbf{z}(0) \sim \mathrm{p}(\cdot; \theta_\mathbf{z}), \ \theta_\mathbf{z} \in \mathbb{R}^{d_{\theta_\mathbf{z}}} \text{ — начальное условие задачи Коши;} \\ \dfrac{\mathrm{d}\,\mathbf{z}(t)}{\mathrm{d}\,t} = f(\mathbf{z}(t), t, \theta); \\ \mathbf{z}(t_i) = \mathbf{z}_{t_i}, \ i = \overline{1, N} \text{ — значения в узлах разностной схемы решения задачи Коши;} \\ \mathbf{x}_{t_i} \sim \mathrm{p}(\cdot|\mathbf{z}_{t_i}; \theta_\mathbf{x}), \ \theta_\mathbf{x} \in \mathbb{R}^{d_{\theta_\mathbf{x}}} \text{ — сэмплирующая модель.} \end{cases}$$

Причём сэмплировать относительно обучающей выборки можно и в «прошлое», по убыванию значения времени, и в «будущее», по возрастанию значения времени, решая соответствующим образом ОДУ.

3.2 Нейронные стохастические дифференциальные уравнения

Наряду с параметризацией нейросетевых моделей с помощью дифференциальных уравнений рассмотрим параметризацию этих моделей с помощью стохастических дифференциальных уравнений. Данные модели применяются при решении задач классификации, в моделировании динамических систем и в оптимальном управлении [23–26]. В работе подобные нейросетевые модели параметризуются в рамках следующей стохастической задачи Коши:

$$\begin{cases} \mathrm{d}\,\mathbf{z}(t) = f(\mathbf{z}(t), t, \theta)\,\mathrm{d}\,t + G(\mathbf{z}(t), t, \theta)\,\mathrm{d}\,W_t; \\ \mathbf{z}(0) = \mathbf{z}_0. \end{cases} \tag{2}$$

В системе $(\mathbf{z}(t), t, \theta) \in \mathbb{R}^{d_{\mathbf{z}}} \times [0, T] \times \mathbb{R}^{\theta}$, $W_t = ({}_1W_t, \ \ldots, \ {}_mW_t)^{\mathrm{T}} \in \mathbb{R}^m$ — m–мерный стандартный винеровский процесс (броуновское движение), обладающий следующими свойствами [11, 27]:

- $W_0 = \mathbf{0}_m$;
- W_t — почти наверно непрерывный;
- W_t имеет независимые приращения;
- $W_t - W_s \sim \mathcal{N}(\mathbf{0}_m, (t-s)I_m), \ 0 \leq s \leq t$.

Стоит заметить, что реализации броуновского движения $W(\omega) = \{W_t = W_t(\omega); t \in [0, T]\}$ имеют почти всюду непрерывные траектории $\omega(\cdot) : \omega(t) = W_t = W_t(\omega), \ t \in [0, T]$, принадлежащие пространству Винера $\mathbb{W} = \mathcal{C}([0, T] : \mathbb{R}^m)$, в силу своих свойств сепарабельности и полноты пространство $\mathbb{W}$ также является польским. То есть реализации броуновского движения можно представить как почти наверно непрерывные отображения $\omega : [0, T] \to \mathbb{R}^m$ [28]. Более того, существует такой единственный вероятностный закон μ для пространства $\mathbb{W}$, называемый винеровской мерой, что:

$$W_0 = \mathbf{0}_m, \ 0 = t_0 < t_1 < \cdots < t_N = T, \ W_{t_{i+1}} - W_{t_i} \sim \mathcal{N}(\mathbf{0}_m, (t_{i+1} - t_i)I_m), \ i = \overline{0, N-1}.$$

Функции

$$f : \mathbb{R}^{d_{\mathbf{z}}} \times \mathbb{R}^{+} \times \mathbb{R}^{d_\theta} \to \mathbb{R}^{d_{\mathbf{z}}}, \ G : \mathbb{R}^{d_{\mathbf{z}}} \times \mathbb{R}^{+} \times \mathbb{R}^{d_\theta} \to \mathbb{R}^{d_{\mathbf{z}} \times m}$$

являются непрерывными по Липшицу и дважды дифференцируемыми. Также предполагается выполнение следующих ограничений на функции $f, \ G$:

1. $\|f(\mathbf{z}, t, \theta)\|_2 + \|G(\mathbf{z}, t, \theta)\|_2 \leq c(1 + \|\mathbf{z}\|_2), \ c > 0, \ \forall (\mathbf{z}, t, \theta) \in \mathbb{R}^{d_{\mathbf{z}}} \times \mathbb{R}^{+} \times \mathbb{R}^{d_\theta}$;
2. $\mathbb{E}_\mu \left[\int\limits_0^T \|f(\mathbf{z}(t), t, \theta) - f(\mathbf{z}(t), t, \theta')\|_2^2 \,\mathrm{d}\,t \right] \leq c_1 \|\theta - \theta'\|_2^2, \ c_1 > 0, \ \forall (\theta, \theta') \in \mathbb{R}^{2d_\theta}, \ \forall \mathbf{z}_0 \in \mathbb{R}^{d_{\mathbf{z}}}$;
3. $\mathbb{E}_\mu \left[\int\limits_0^T \|G(\mathbf{z}(t), t, \theta) - G(\mathbf{z}(t), t, \theta')\|_2^2 \,\mathrm{d}\,t \right] \leq c_2 \|\theta - \theta'\|_2^2, \ c_2 > 0, \ \forall (\theta, \theta') \in \mathbb{R}^{2d_\theta}, \ \forall \mathbf{z}_0 \in \mathbb{R}^{d_{\mathbf{z}}}$.

(3)

Выполнение первого из условий в (3) гарантирует устойчивость по Ляпунову решения стохастической задачи Коши (2) [26]. Будем называть ограничения в (3) *условиями регулярности*. В свою очередь решение задачи (2) называется случайным *процессом диффузии*. Стоит также заметить, что в данной работе рассматриваются только *борелевские отображения*.

3.2.1 Настройка параметров модели

Для вывода соответствующих формул рассмотрим следующую дважды дифференцируемую липшицевую функцию потерь $l(\cdot)$ и её среднее значение L:

$$L = \mathbb{E}_\mu \left[l(\mathbf{z}(T)) \right] = \mathbb{E}_\mu \left[l \left(\mathbf{z}_0 + \int_0^T f(\mathbf{z}(t), t, \theta) \, \mathrm{d}\, t + \int_0^T G(\mathbf{z}(t), t, \theta) \, \mathrm{d}\, W_t \right) \right]. \tag{4}$$

Для настройки параметров $\theta \in \mathbb{R}^{d_\theta}$ на практике часто используются градиентные методы, для которых средний градиент оценивается с помощью методов Монте–Карло [29], используя *path–wise* метод [23, 25]. Подробнее вывод необходимых градиентов для оптимизации модели, заданной (2), рассмотрен в разделе 4.

3.2.2 Моделирование решения стохастического дифференциального уравнения

Прежде чем перейти непосредственно к описанию разностной схемы, с помощью которой в общем случае моделируют процесс диффузии, следует указать, что стохастические дифференциальные уравнения (СДУ) рассматриваются в рамках теории интеграла Ито [27]. СДУ определены на полном каноническом вероятностном пространстве $(\Omega, \mathcal{F}, \mu)$,

$$\omega \in \Omega = \mathcal{C}([0, T] : \mathbb{R}^m), \ \mathcal{F} = \mathcal{B}(\mathcal{C}([0, T] : \mathbb{R}^m)), \ \mu - m\text{–мерная винеровская мера},$$

с упорядоченной последовательностью борелевских σ–алгебр $\mathbb{F} = \{\mathcal{F}_t\}_{t \in [0,T]}$,
$\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}, \ s \leq t$:

$$\mathcal{F}_t := \sigma(\sigma(\{W_s; 0 \leq s \leq t\}) \cup \{\mathcal{O} \subseteq \Omega : \ \exists B \in \mathcal{F}, \ \mathcal{O} \subseteq B, \ \mu(B) = 0\}).$$

Ограничения на функции $f, \ G$ такие же, как в случае (2), и заданы (3). Для вывода рассматриваемой разностной схемы выведем следующие утверждения.

Лемма 1 (Обобщённая формула Ито, [30]). *Пусть*

$$\mathrm{d}\, \mathbf{z}(t) = f(\mathbf{z}(t), t, \theta) \, \mathrm{d}\, t + G(\mathbf{z}(t), t, \theta) \, \mathrm{d}\, W_t$$

является $d_{\mathbf{z}}$–мерным процессом Ито. Пусть $g(x, t) = (g_1(x, t), \ \ldots, \ g_p(x, t))$ — дважды непрерывно дифференцируемое отображение из $\mathbb{R}^{d_{\mathbf{z}}} \times [0, \infty)$ в $\mathbb{R}^p$. Тогда процесс

$$Y(t) = \widetilde{Y}(t, \omega) = g(\mathbf{z}(t), t)$$

является процессом Ито, который покомпонентно задаётся в виде следующих соотношений:

$$\mathrm{d}\, Y_k(t) = \frac{\partial g_k}{\partial t}(\mathbf{z}(t), t)\,\mathrm{d}\, t + \sum_{i=1}^{d_{\mathbf{z}}} \frac{\partial g_k}{\partial x_i}(\mathbf{z}(t), t)\,\mathrm{d}\, \mathbf{z}_i(t) + \frac{1}{2}\sum_{i,j=1}^{d_{\mathbf{z}}} \frac{\partial^2 g_k}{\partial x_i \partial x_j}(\mathbf{z}(t), t)\,\mathrm{d}\, \mathbf{z}_i(t)\,\mathrm{d}\, \mathbf{z}_j(t),$$

где $\mathrm{d}_i W_t\,\mathrm{d}_j W_t = \delta_{ij}\,\mathrm{d}\, t,\ (\mathrm{d}\, t)^2 = \mathrm{d}_i W_t\,\mathrm{d}\, t = \mathrm{d}\, t\,\mathrm{d}_i W_t = 0,\ \delta_{ij}$ *— символ Кронекера.*

Представленное выше утверждение переформулировано в более удобном виде для дальнейших выводов в лемме 2. Но, прежде чем перейти к лемме 2, из соображений удобства обозначим соответствующие первые частные производные

$$\left(\frac{\partial g_i(\mathbf{z}(t), t)}{\partial x_j}\right)_{i,j=1}^{p,d_{\mathbf{z}}}$$

вектор–функции $g(x, t)$ как

$$\frac{\partial g}{\partial x}(\mathbf{z}(t), t) = \frac{\partial g(\mathbf{z}(t), t)}{\partial x} = \frac{\partial g(\mathbf{z}(t), t)}{\partial \mathbf{z}} = \frac{\partial g(\mathbf{z}(t), t)}{\partial \mathbf{z}(t)},$$

для функции–матрицы $g(x, t) = (g_{kl}(x, t))_{k,l=1}^{p,q}$ нотация для частной производной

$$\left(\frac{\partial g_{kl}(\mathbf{z}(t), t)}{\partial x_i}\right)_{i=1}^{d_{\mathbf{z}}}$$

следующая:

$$\frac{\partial g_{kl}}{\partial x}(\mathbf{z}(t), t) = \frac{\partial (e_k^{\mathrm{T}} g(\mathbf{z}(t), t) e_l)}{\partial x} = \frac{\partial (e_k^{\mathrm{T}} g(\mathbf{z}(t), t) e_l)}{\partial \mathbf{z}} = \frac{\partial (e_k^{\mathrm{T}} g(\mathbf{z}(t), t) e_l)}{\partial \mathbf{z}(t)},\ k = \overline{1, p},\ l = \overline{1, q};$$

$e_k,\ e_l$ — k–ый и l–ый орты. Вторые частные производные

$$\left(\frac{\partial^2 g_{kl}(\mathbf{z}(t), t)}{\partial x_i \partial x_j}\right)_{i,j=1}^{d_{\mathbf{z}},d_{\mathbf{z}}}$$

функции–матрицы $g(x, t) = (g_{kl}(x, t))_{k,l=1}^{p,q}$ обозначим как

$$\frac{\partial^2 g_{kl}}{\partial x \partial x^{\mathrm{T}}}(\mathbf{z}(t), t) = \frac{\partial^2 (e_k^{\mathrm{T}} g(\mathbf{z}(t), t) e_l)}{\partial x \partial x^{\mathrm{T}}} = \frac{\partial^2 (e_k^{\mathrm{T}} g(\mathbf{z}(t), t) e_l)}{\partial \mathbf{z} \partial \mathbf{z}^{\mathrm{T}}} = \frac{\partial^2 (e_k^{\mathrm{T}} g(\mathbf{z}(t), t) e_l)}{\partial \mathbf{z}(t) \partial \mathbf{z}(t)^{\mathrm{T}}},\ k = \overline{1, p},\ l = \overline{1, q}.$$

Вторые частные производные

$$\left(\frac{\partial^2 g_k(\mathbf{z}(t), t)}{\partial x_i \partial x_j}\right)_{i,j=1}^{d_{\mathbf{z}},d_{\mathbf{z}}}$$

функции–вектора $g(x, t) = (g_k(x, t))_{k=1}^{p}$ можно рассмотреть как упрощение случая функции–

матрицы:

$$\frac{\partial^2 g_k}{\partial x \partial x^{\mathrm{T}}}(\mathbf{z}(t),t) = \frac{\partial^2 (e_k^{\mathrm{T}} g(\mathbf{z}(t),t))}{\partial x \partial x^{\mathrm{T}}} = \frac{\partial^2 (e_k^{\mathrm{T}} g(\mathbf{z}(t),t))}{\partial \mathbf{z} \partial \mathbf{z}^{\mathrm{T}}} = \frac{\partial^2 (e_k^{\mathrm{T}} g(\mathbf{z}(t),t))}{\partial \mathbf{z}(t) \partial \mathbf{z}(t)^{\mathrm{T}}},\ k = \overline{1,p}.$$

Лемма 2. *Пусть*

$$\mathrm{d}\,\mathbf{z}(t) = f(\mathbf{z}(t),t,\theta)\,\mathrm{d}\,t + G(\mathbf{z}(t),t,\theta)\,\mathrm{d}\,W_t$$

является D–мерным процессом Ито. Пусть $g(x,t) = (g_i(x,t))_{i=1}^p$ — дважды непрерывно дифференцируемое отображение из $\mathbb{R}^{d_{\mathbf{z}}} \times [0,\infty)$ в $\mathbb{R}^p$. Тогда процесс

$$Y(t) = \widetilde{Y}(t,\omega) = g(\mathbf{z}(t),t)$$

является процессом Ито, который задаётся в виде следующих соотношений:

$$\begin{aligned}\mathrm{d}\,Y(t) = \Bigg(&\frac{\partial g}{\partial t}(\mathbf{z}(t),t) + \frac{\partial g(\mathbf{z}(t),t)}{\partial x} f(\mathbf{z}(t),t,\theta) + \\ &+ \frac{1}{2}\sum_{i=1}^{p} \mathrm{tr}\left(G^{\mathrm{T}}(\mathbf{z}(t),t,\theta)\frac{\partial^2 (e_i^{\mathrm{T}} g(\mathbf{z}(t),t))}{\partial x \partial x^{\mathrm{T}}} G(\mathbf{z}(t),t,\theta)\right) e_i\Bigg)\,\mathrm{d}\,t + \\ &+ \frac{\partial g(\mathbf{z}(t),t)}{\partial x} G(\mathbf{z}(t),t,\theta)\,\mathrm{d}\,W_t,\end{aligned}$$

где e_i — i–ый орт в $\mathbb{R}^p$. Если $g(x,t) = (g_{ij}(x,t))_{i,j=1}^{p,q}$ — дважды непрерывно дифференцируемое отображение из $\mathbb{R}^{d_{\mathbf{z}}} \times [0,\infty)$ в $\mathbb{R}^{p\times q}$. Тогда процесс

$$Y(t) = \widetilde{Y}(t,\omega) = g(\mathbf{z}(t),t)$$

является процессом Ито, который задаётся в виде следующих соотношений:

$$\begin{aligned}\mathrm{d}\,Y(t) = &\frac{\partial g}{\partial t}(\mathbf{z}(t),t)\,\mathrm{d}\,t + \\ &+ \sum_{k,l=1}^{p,q}\left(\frac{\partial (e_k^{\mathrm{T}} g(\mathbf{z}(t),t) e_l)}{\partial x} f(\mathbf{z}(t),t,\theta)\right) e_k e_l^{\mathrm{T}}\,\mathrm{d}\,t + \\ &+ \frac{1}{2}\sum_{k,l=1}^{p,q}\mathrm{tr}\left(G^{\mathrm{T}}(\mathbf{z}(t),t,\theta)\frac{\partial^2 (e_k^{\mathrm{T}} g(\mathbf{z}(t),t) e_l)}{\partial x \partial x^{\mathrm{T}}} G(\mathbf{z}(t),t,\theta)\right) e_k e_l^{\mathrm{T}}\,\mathrm{d}\,t + \\ &+ \sum_{k,l=1}^{p,q} e_k e_l^{\mathrm{T}}\left(\frac{\partial (e_k^{\mathrm{T}} g(\mathbf{z}(t),t) e_l)}{\partial x} G(\mathbf{z}(t),t,\theta)\,\mathrm{d}\,W_t\right),\end{aligned}$$

где e_k — k–ый орт в $\mathbb{R}^p$, e_l — l–ый орт в $\mathbb{R}^q$.

Доказательство. Подставляя в формулу для $\mathrm{d}\,Y_k$ из леммы 1 выражение для элемента вектора

$$\mathrm{d}\,\mathbf{z}_i(t) = f_i(\mathbf{z}(t),t,\theta)\,\mathrm{d}\,t + \sum_{j=1}^{m} G_{ij}(\mathbf{z}(t),t,\theta)\,\mathrm{d}_j W_t,$$

выполним последовательность преобразований, используя свойства процессов Ито из леммы 1. Сначала преобразуем произведение $\mathrm{d}\,\mathbf{z}_i(t)\,\mathrm{d}\,\mathbf{z}_j(t)$:

$$\begin{aligned}
&\mathrm{d}\,\mathbf{z}_i(t)\,\mathrm{d}\,\mathbf{z}_j(t) = \\
&= \left(f_i(\mathbf{z}(t),t,\theta)\,\mathrm{d}\,t + \sum_{k=1}^{m} G_{ik}(\mathbf{z}(t),t,\theta)\,\mathrm{d}_k W_t\right)\left(f_j(\mathbf{z}(t),t,\theta)\,\mathrm{d}\,t + \sum_{l=1}^{m} G_{jl}(\mathbf{z}(t),t,\theta)\,\mathrm{d}_l W_t\right) = \\
&= f_i(\mathbf{z}(t),t,\theta) f_j(\mathbf{z}(t),t,\theta)\underbrace{(\mathrm{d}\,t)^2}_{=0} + \sum_{l=1}^{m} f_i(\mathbf{z}(t),t,\theta) G_{jl}(\mathbf{z}(t),t,\theta)\underbrace{\mathrm{d}\,t\,\mathrm{d}_l W_t}_{=0} + \\
&+ \sum_{k=1}^{m} G_{ik}(\mathbf{z}(t),t,\theta) f_j(\mathbf{z}(t),t,\theta)\underbrace{\mathrm{d}_k W_t\,\mathrm{d}\,t}_{=0} + \sum_{k,l=1}^{m} G_{ik}(\mathbf{z}(t),t,\theta) G_{jl}(\mathbf{z}(t),t,\theta)\underbrace{\mathrm{d}_k W_t\,\mathrm{d}_l W_t}_{=\delta_{kl}\,\mathrm{d}\,t} = \\
&= \sum_{k=1}^{m} G_{ik}(\mathbf{z}(t),t) G_{jk}(\mathbf{z}(t),t,\theta)\,\mathrm{d}\,t.
\end{aligned} \tag{5}$$

Подставим выражения (5) и $\mathrm{d}\,\mathbf{z}_i(t)$ в $\mathrm{d}\,Y_k(t)$:

$$\begin{aligned}
\mathrm{d}\,Y_k(t) = \frac{\partial g_k}{\partial t}(\mathbf{z}(t),t)\,\mathrm{d}\,t + \sum_{i=1}^{d_{\mathbf{z}}} \frac{\partial g_k}{\partial x_i}(\mathbf{z}(t),t)\left(f_i(\mathbf{z}(t),t,\theta)\,\mathrm{d}\,t + \sum_{j=1}^{m} G_{ij}(\mathbf{z}(t),t,\theta)\,\mathrm{d}_j W_t\right) + \\
+ \frac{1}{2}\sum_{i,j,l=1}^{d_{\mathbf{z}},d_{\mathbf{z}},m} \frac{\partial^2 g_k}{\partial x_i \partial x_j}(\mathbf{z}(t),t) G_{il}(\mathbf{z}(t),t,\theta) G_{jl}(\mathbf{z}(t),t,\theta)\,\mathrm{d}\,t = \left(\frac{\partial g_k}{\partial t}(\mathbf{z}(t),t) + \right. \\
\left. + \sum_{i=1}^{d_{\mathbf{z}}} \frac{\partial g_k}{\partial x_i}(\mathbf{z}(t),t) f_i(\mathbf{z}(t),t,\theta) + \frac{1}{2}\sum_{i,j,l=1}^{d_{\mathbf{z}},d_{\mathbf{z}},m} \frac{\partial^2 g_k}{\partial x_i \partial x_j}(\mathbf{z}(t),t) G_{il}(\mathbf{z}(t),t,\theta) G_{jl}(\mathbf{z}(t),t,\theta)\right)\mathrm{d}\,t + \\
+ \sum_{i,j=1}^{d_{\mathbf{z}},m} \frac{\partial g_k}{\partial x_i}(\mathbf{z}(t),t) G_{ij}(\mathbf{z}(t),t,\theta)\,\mathrm{d}_j W_t.
\end{aligned} \tag{6}$$

Переписав (6) в матричном виде, получим следующее выражение для вектор–функций:

$$\begin{aligned}
\mathrm{d}\,Y(t) = \left(\frac{\partial g}{\partial t}(\mathbf{z}(t),t) + \frac{\partial g}{\partial x}(\mathbf{z}(t),t) f(\mathbf{z}(t),t,\theta) + \right. \\
\left. + \frac{1}{2}\sum_{i=1}^{p} \mathrm{tr}\left(G^{\mathrm{T}}(\mathbf{z}(t),t,\theta)\frac{\partial^2 (e_i^{\mathrm{T}} g(\mathbf{z}(t),t))}{\partial x \partial x^{\mathrm{T}}} G(\mathbf{z}(t),t,\theta)\right) e_i\right)\mathrm{d}\,t + \\
+ \frac{\partial g(\mathbf{z}(t),t)}{\partial x} G(\mathbf{z}(t),t,\theta)\,\mathrm{d}\,W_t.
\end{aligned}$$

Соответственно, аналогичным образом выводится следующая формула в случае $(g_{ij}(x,t))_{i,j=1}^{p,q}$:

$$\begin{aligned}\mathrm{d}\,Y(t) = \frac{\partial g}{\partial t}(\mathbf{z}(t),t)\,\mathrm{d}\,t + \sum_{k,l=1}^{p,q}\left(\frac{\partial (e_k^{\mathrm{T}} g(\mathbf{z}(t),t)e_l)}{\partial x} f(\mathbf{z}(t),t,\theta)\right) e_k e_l^{\mathrm{T}}\,\mathrm{d}\,t + \\ + \frac{1}{2}\sum_{k,l=1}^{p,q}\mathrm{tr}\left(G^{\mathrm{T}}(\mathbf{z}(t),t,\theta)\frac{\partial^2 (e_k^{\mathrm{T}} g(\mathbf{z}(t),t)e_l)}{\partial x \partial x^{\mathrm{T}}} G(\mathbf{z}(t),t,\theta)\right) e_k e_l^{\mathrm{T}}\,\mathrm{d}\,t + \\ + \sum_{k,l=1}^{p,q} e_k e_l^{\mathrm{T}}\left(\frac{\partial (e_k^{\mathrm{T}} g(\mathbf{z}(t),t)e_l)}{\partial x} G(\mathbf{z}(t),t,\theta)\,\mathrm{d}\,W_t\right).\end{aligned} \tag{7}$$

То есть для вывода (7) достаточно заменить в (6) индекс элемента вектора k на индексы элемента массива (k,l) и собрать поэлементно матрицы с помощью орт $e_k e_l^{\mathrm{T}}$.

□

Лемма 2 позволяет переписать дифференциал процесса Ито $Y(t) = g(\mathbf{z}(t),t)$ с помощью дифференциальных операторов $\mathcal{D}^0,\ \mathcal{D}^1$:

$$\mathrm{d}\,Y(t) = \mathcal{D}^0(Y(t))\,\mathrm{d}\,t + \mathcal{D}^1(Y(t),\mathrm{d}\,W_t).$$

В случае вектор–функции $g(\mathbf{z}(t),t)$:

- $\mathcal{D}^0(g(\mathbf{z}(t),t)) = \dfrac{\partial g}{\partial t}(\mathbf{z}(t),t) + \dfrac{\partial g(\mathbf{z}(t),t)}{\partial \mathbf{z}} f(\mathbf{z}(t),t,\theta) + \\ + \dfrac{1}{2}\sum\limits_{i=1}^{p}\mathrm{tr}\left(G^{\mathrm{T}}(\mathbf{z}(t),t,\theta)\dfrac{\partial^2 (e_i^{\mathrm{T}} g(\mathbf{z}(t),t))}{\partial \mathbf{z}\partial \mathbf{z}^{\mathrm{T}}} G(\mathbf{z}(t),t,\theta)\right) e_i;$
- $\mathcal{D}^1(g(\mathbf{z}(t),t),\mathrm{d}\,W_t) = \dfrac{\partial g(\mathbf{z}(t),t)}{\partial \mathbf{z}} G(\mathbf{z}(t),t,\theta)\,\mathrm{d}\,W_t.$

Для функции–матрицы $g(\mathbf{z}(t),t)$ операторы имеют следующее представление:

- $\mathcal{D}^0(g(\mathbf{z}(t),t)) = \dfrac{\partial g}{\partial t}(\mathbf{z}(t),t) + \sum\limits_{k,l=1}^{p,q}\left(\underbrace{\dfrac{\partial (e_k^{\mathrm{T}} g(\mathbf{z}(t),t)e_l)}{\partial \mathbf{z}} f(\mathbf{z}(t),t,\theta)}_{\text{скаляр}}\right) e_k e_l^{\mathrm{T}} + \\ + \dfrac{1}{2}\sum\limits_{k,l=1}^{p,q}\mathrm{tr}\left(G^{\mathrm{T}}(\mathbf{z}(t),t,\theta)\dfrac{\partial^2 (e_k^{\mathrm{T}} g(\mathbf{z}(t),t)e_l)}{\partial \mathbf{z}\partial \mathbf{z}^{\mathrm{T}}} G(\mathbf{z}(t),t,\theta)\right) e_k e_l^{\mathrm{T}};$
- $\mathcal{D}^1(g(\mathbf{z}(t),t),\mathrm{d}\,W_t) = \sum\limits_{k,l=1}^{p,q} e_k e_l^{\mathrm{T}}\left(\underbrace{\dfrac{\partial (e_k^{\mathrm{T}} g(\mathbf{z}(t),t)e_l)}{\partial \mathbf{z}} G(\mathbf{z}(t),t,\theta)\,\mathrm{d}\,W_t}_{\text{скаляр}}\right).$

Таким образом, можно с помощью данных операторов переписать решение стохастической задачи Коши (8):

$$\begin{cases} \mathrm{d}\,\mathbf{z}(t) = f(\mathbf{z}(t), t, \theta)\,\mathrm{d}\,t + G(\mathbf{z}(t), t, \theta)\,\mathrm{d}\,W_t; \\ \mathbf{z}(t_0) = \mathbf{z}_0,\ (t_0, t) \in [0, T]^2,\ t_0 \leq t; \end{cases} \tag{8}$$

$$\mathbf{z}(t) = \mathbf{z}(t_0) + \int\limits_{t_0}^{t} \mathcal{D}^0(\mathbf{z}(s))\,\mathrm{d}\,s + \int\limits_{t_0}^{t} \mathcal{D}^1(\mathbf{z}(s), \mathrm{d}\,W_s).$$

Более того, данное представление справедливо и для функций $f(\mathbf{z}(t), t, \theta),\ G(\mathbf{z}(t), t, \theta)$ при условии дважды непрерывной дифференцируемости (лемма 2), $(t_0, t) \in [0, T]^2,\ t_0 \leq t$:

$$\begin{cases} f(\mathbf{z}(t), t, \theta) = f(\mathbf{z}(t_0), t_0, \theta) + \int\limits_{t_0}^{t} \mathcal{D}^0(f(\mathbf{z}(s), s, \theta))\,\mathrm{d}\,s + \int\limits_{t_0}^{t} \mathcal{D}^1(f(\mathbf{z}(s), s, \theta), \mathrm{d}\,W_s); \\ G(\mathbf{z}(t), t, \theta) = G(\mathbf{z}(t_0), t_0, \theta) + \int\limits_{t_0}^{t} \mathcal{D}^0(G(\mathbf{z}(s), s, \theta))\,\mathrm{d}\,s + \int\limits_{t_0}^{t} \mathcal{D}^1(G(\mathbf{z}(s), s, \theta), \mathrm{d}\,W_s). \end{cases} \tag{9}$$

Полученное выражение решения (8) может быть представлено с помощью (9) способом, указанным в лемме 3.

Лемма 3. *Пусть*

$$\mathrm{d}\,\mathbf{z}(t) = f(\mathbf{z}(t), t, \theta)\,\mathrm{d}\,t + G(\mathbf{z}(t), t, \theta)\,\mathrm{d}\,W_t$$

является $d_{\mathbf{z}}$–мерным процессом Ито. Функция f — дважды непрерывно дифференцируемая, G — дважды непрерывно дифференцируемая функция, обе функции непрерывны по Липшицу. Тогда $\mathbf{z}(t)$ почти наверно имеет следующее представление при $(t_0, t) \in [0, T]^2,\ t_0 \leq t$:

$$\mathbf{z}(t) = \mathbf{z}(t_0) + f(\mathbf{z}(t_0), t_0, \theta)(t - t_0) + G(\mathbf{z}(t_0), t_0, \theta)(W_t - W_{t_0}) + R_1(t_0, t, \theta). \tag{10}$$

Доказательство. Доказательство заключается в последовательном применении операторов $\mathcal{D}^0, \mathcal{D}^1$:

$$\begin{aligned} \mathbf{z}(t) = \mathbf{z}(t_0) + \int\limits_{t_0}^{t} f(\mathbf{z}(s), s, \theta)\,\mathrm{d}\,s + \int\limits_{t_0}^{t} G(\mathbf{z}(s), s, \theta)\,\mathrm{d}\,W_s = \mathbf{z}(t_0) + \\ + \int\limits_{t_0}^{t} \left(f(\mathbf{z}(t_0), t_0, \theta) + \int\limits_{t_0}^{s} \mathcal{D}^0(f(\mathbf{z}(s_1), s_1, \theta))\,\mathrm{d}\,s_1 + \int\limits_{t_0}^{s} \mathcal{D}^1(f(\mathbf{z}(s_1), s_1, \theta), \mathrm{d}\,W_{s_1}) \right) \mathrm{d}\,s + \\ + \int\limits_{t_0}^{t} \left(G(\mathbf{z}(t_0), t_0, \theta) + \int\limits_{t_0}^{s} \mathcal{D}^0(G(\mathbf{z}(s_1), s_1, \theta))\,\mathrm{d}\,s_1 + \int\limits_{t_0}^{s} \mathcal{D}^1(G(\mathbf{z}(s_1), s_1, \theta), \mathrm{d}\,W_{s_1}) \right) \mathrm{d}\,W_s. \end{aligned} \tag{11}$$

Перепишем выражение (11):

$$\begin{aligned}\mathbf{z}(t) = \mathbf{z}(t_0) + f(\mathbf{z}(t_0), t_0, \theta)(t - t_0) + G(\mathbf{z}(t_0), t_0, \theta)(W_t - W_{t_0}) + \\ + \int_{t_0}^{t} \left(\int_{t_0}^{s} \mathcal{D}^0(f(\mathbf{z}(s_1), s_1, \theta))\,\mathrm{d}\,s_1 + \int_{t_0}^{s} \mathcal{D}^1(f(\mathbf{z}(s_1), s_1, \theta), \mathrm{d}\,W_{s_1}) \right) \mathrm{d}\,s + \\ + \int_{t_0}^{t} \left(\int_{t_0}^{s} \mathcal{D}^0(G(\mathbf{z}(s_1), s_1, \theta))\,\mathrm{d}\,s_1 + \int_{t_0}^{s} \mathcal{D}^1(G(\mathbf{z}(s_1), s_1, \theta), \mathrm{d}\,W_{s_1}) \right) \mathrm{d}\,W_s.\end{aligned} \tag{12}$$

Обозначив за $R_1(t_0, t, \theta)$ нижние две строки в (12), доказываем требуемое. □

Полученное представление в (10) без остаточного члена $R_1(t_0, t, \theta)$ представляет собой итеративную формулу, используемую в *методе Эйлера–Маруямы* моделирования случайного процесса $\mathbf{z}(t)$ на разностной схеме [31], с помощью которого в данной работе осуществлено моделирование процессов диффузии $\mathbf{z}(t)$. Алгоритмическое описание метода Эйлера–Маруямы представлено в листинге 1, процедура, осуществляющая моделирование $\mathbf{z}(t)$ называется SDESolve.

Algorithm 1 Сэмплирование реализации решения стохастической задачи Коши на разностной схеме методом Эйлера–Маруямы.

function SDESolve((f, θ), (G, θ), $\mathbf{z}_{t_0}$, $(t_0, \ \ldots, \ t_N)$) ▷ $\mathbf{z}_{t_0}$ — начальная точка, t_0 — момент начала моделирования

  $\mathrm{I} = (\mathbf{z}_{t_0})$ ▷ I — реализация процесса диффузии

  for $i = 1, 2, \ \ldots, \ N$ **do**

    $\mathbf{z}_{t_i} = \mathbf{z}_{t_{i-1}} + f(\mathbf{z}(t_{i-1}), t_{i-1}, \theta)(t_i - t_{i-1}) + G(\mathbf{z}(t_{i-1}), t_{i-1}, \theta)(W_{t_i} - W_{t_{i-1}})$

    Добавить сэмпл $\mathbf{z}_{t_i}$ в I

  return I ▷ $\mathrm{I} = (\mathbf{z}_{t_0}, \ \ldots, \ \mathbf{z}_{t_N})$

3.3 Вариационный вывод с дивергенциями α–Реньи

Рассмотрим вероятностную модель с параметром θ для наблюдаемой переменной $\mathbf{x}$ и её скрытой (латентной) переменной $\mathbf{z}$:

$$\mathrm{p}(\mathbf{x}, \mathbf{z}; \theta) = \mathrm{p}(\mathbf{x}|\mathbf{z}; \theta)\,\mathrm{p}(\mathbf{z}; \theta). \tag{13}$$

Вне зависимости от поставленной задачи, будь то вычисление маргинального распределения $\mathrm{p}(\mathbf{x}|\theta)$ или апостериорного распределения $\mathrm{p}(\mathbf{z}|\mathbf{x}; \theta)$, основная трудность заключается в вычислении следующего интеграла, называемым *обоснованностью*:

$$\mathrm{p}(\mathbf{x}; \theta) = \mathbb{E}_{\mathrm{p}(\mathbf{z};\theta)} \left[\mathrm{p}(\mathbf{x}|\mathbf{z}; \theta) \right]. \tag{14}$$

В случаях, когда вычисление (14) является практически невозможным, например, в моделях процессов диффузии, для вывода в модели (13) используется процедура вариационного вывода [32, 33]. Процедура вариационного вывода заключается в приближении апостериорного распределения другим распределением $\mathrm{q}(\mathbf{z};\phi)$ с параметром ϕ, называемым *вариационным приближением*, что позволяет практически снизу оценить (14):

$$
\begin{aligned}
&\mathrm{p}(\mathbf{z}|\mathbf{x};\theta) = \frac{\mathrm{p}(\mathbf{x}|\mathbf{z};\theta)\,\mathrm{p}(\mathbf{z};\theta)}{\mathrm{p}(\mathbf{x};\theta)};\\
&\ln \mathrm{p}(\mathbf{x};\theta) = \mathbb{E}_{\mathrm{q}}\left[\ln \mathrm{p}(\mathbf{x};\theta)\right] = \mathbb{E}_{\mathrm{q}}\left[\ln\left(\frac{\mathrm{p}(\mathbf{x}|\mathbf{z};\theta)\,\mathrm{p}(\mathbf{z};\theta)\,\mathrm{q}(\mathbf{z};\phi)}{\mathrm{q}(\mathbf{z};\phi)\,\mathrm{p}(\mathbf{z}|\mathbf{x};\theta)}\right)\right] = \mathbb{E}_{\mathrm{q}}\left[\ln\left(\frac{\mathrm{q}(\mathbf{z};\phi)}{\mathrm{p}(\mathbf{z}|\mathbf{x};\theta)}\right)\right] + \\
&\quad + \mathbb{E}_{\mathrm{q}}\left[\ln\left(\frac{\mathrm{p}(\mathbf{x}|\mathbf{z};\theta)\,\mathrm{p}(\mathbf{z};\theta)}{\mathrm{q}(\mathbf{z};\phi)}\right)\right] = \underbrace{\mathrm{KL}[\mathrm{q}(\mathbf{z};\phi)||\,\mathrm{p}(\mathbf{z}|\mathbf{x};\theta)]}_{\text{KL–дивергенция, }\geq 0} + \mathbb{E}_{\mathrm{q}}\left[\ln\left(\frac{\mathrm{p}(\mathbf{x}|\mathbf{z};\theta)\,\mathrm{p}(\mathbf{z};\theta)}{\mathrm{q}(\mathbf{z};\phi)}\right)\right] \geq \\
&\quad \geq \mathbb{E}_{\mathrm{q}}\left[\ln\left(\frac{\mathrm{p}(\mathbf{x}|\mathbf{z};\theta)\,\mathrm{p}(\mathbf{z};\theta)}{\mathrm{q}(\mathbf{z};\phi)}\right)\right].
\end{aligned}
\tag{15}
$$

Однако данный способ не является единственным при оценивании логарифма обоснованности, полученная оценка (15) обобщается с помощью дивергенции α–Реньи — $\mathrm{D}_\alpha[\cdot||\cdot]$ [34, 35]:

$$
\begin{aligned}
&\mathrm{D}_\alpha[\mathrm{q}(\mathbf{z};\phi)||\,\mathrm{p}(\mathbf{z}|\mathbf{x};\theta)] = \frac{1}{\alpha-1}\ln \mathbb{E}_{\mathrm{p}(\mathbf{z}|\mathbf{x};\theta)}\left[\left(\frac{\mathrm{q}(\mathbf{z};\phi)}{\mathrm{p}(\mathbf{z}|\mathbf{x};\theta)}\right)^{\alpha}\right],\ \alpha \geq 0;\\
&\quad \ln \mathrm{p}(\mathbf{x};\theta) \geq \ln \mathrm{p}(\mathbf{x};\theta) - \underbrace{\mathrm{D}_\alpha[\mathrm{q}(\mathbf{z};\phi)||\,\mathrm{p}(\mathbf{z}|\mathbf{x};\theta)]}_{\geq 0} = \\
&\quad\quad = \frac{1-\alpha}{1-\alpha}\ln \mathrm{p}(\mathbf{x};\theta) + \frac{1}{1-\alpha}\ln \mathbb{E}_{\mathrm{q}}\left[\left(\frac{\mathrm{p}(\mathbf{z}|\mathbf{x};\theta)}{\mathrm{q}(\mathbf{z};\phi)}\right)^{1-\alpha}\right] = \\
&\quad\quad = \frac{1}{1-\alpha}\ln \mathbb{E}_{\mathrm{q}}\left[\left(\frac{\mathrm{p}(\mathbf{x};\theta)\,\mathrm{p}(\mathbf{z}|\mathbf{x};\theta)}{\mathrm{q}(\mathbf{z};\phi)}\right)^{1-\alpha}\right] = \\
&\quad\quad = \frac{1}{1-\alpha}\ln \mathbb{E}_{\mathrm{q}}\left[\left(\frac{\mathrm{p}(\mathbf{x}|\mathbf{z};\theta)\,\mathrm{p}(\mathbf{z};\theta)}{\mathrm{q}(\mathbf{z};\phi)}\right)^{1-\alpha}\right] = \\
&\quad\quad = \frac{1}{1-\alpha}\ln \mathbb{E}_{\mathrm{q}(\mathbf{z}^{1:K};\phi)}\left[\frac{1}{K}\sum_{k=1}^{K}\left(\frac{\mathrm{p}(\mathbf{x}|\mathbf{z}^k;\theta)\,\mathrm{p}(\mathbf{z}^k;\theta)}{\mathrm{q}(\mathbf{z}^k;\phi)}\right)^{1-\alpha}\right] \geq \\
&\quad\quad \geq \{\text{неравенство Йенсена для вогнутой } \ln(\cdot)\} \geq \\
&\quad\quad \geq \mathcal{L}_\alpha^K(\mathbf{x};\phi,\theta) = \frac{1}{1-\alpha}\mathbb{E}_{\mathrm{q}(\mathbf{z}^{1:K};\phi)}\left[\ln\left(\frac{1}{K}\sum_{k=1}^{K}\left(\frac{\mathrm{p}(\mathbf{x}|\mathbf{z}^k;\theta)\,\mathrm{p}(\mathbf{z}^k;\theta)}{\mathrm{q}(\mathbf{z}^k;\phi)}\right)^{1-\alpha}\right)\right],\\
&\quad\quad \mathbf{z}^k \overset{\text{i.i.d.}}{\sim} \mathrm{q}(\cdot;\phi),\ \mathrm{q}(\mathbf{z}^{1:K};\phi) = \prod_{k=1}^{K}\mathrm{q}(\mathbf{z}^k;\phi).
\end{aligned}
\tag{16}
$$

Полученная в (16) *вариационная нижняя оценка* $\mathcal{L}_\alpha^K(\mathbf{x};\phi,\theta)$ на $\ln p(\mathbf{x};\theta)$ позволяет непрерывно объединить оценки на логарифм обоснованности, перечисленные в таблице 1.

Значение α	**Нижняя оценка** $\ln \mathrm{p}(\mathbf{x};\theta)$	**Название**
0	$\mathbb{E}_{\mathrm{q}(\mathbf{z}^{1:K};\phi)}\left[\ln\left(\frac{1}{K}\sum\limits_{k=1}^{K}\left(\frac{\mathrm{p}(\mathbf{x}\|\mathbf{z}^k;\theta)\,\mathrm{p}(\mathbf{z}^k;\theta)}{\mathrm{q}(\mathbf{z}^k;\phi)}\right)\right)\right]$	оценка $\ln \mathrm{p}(\mathbf{x};\theta)$ с помощью выборки по значимости [36]
1	$\mathbb{E}_{\mathrm{q}(\mathbf{z}^{1:K};\phi)}\left[\frac{1}{K}\sum\limits_{k=1}^{K}\ln\left(\frac{\mathrm{p}(\mathbf{x}\|\mathbf{z}^k;\theta)\,\mathrm{p}(\mathbf{z}^k;\theta)}{\mathrm{q}(\mathbf{z}^k;\phi)}\right)\right]$	вариационный вывод с KL–дивергенцией [32]
$+\infty$	$\mathbb{E}_{\mathrm{q}(\mathbf{z}^{1:K};\phi)}\left[\min\limits_{k\in\{1,\ldots,K\}}\ln\left(\frac{\mathrm{p}(\mathbf{x}\|\mathbf{z}^k;\theta)\,\mathrm{p}(\mathbf{z}^k;\theta)}{\mathrm{q}(\mathbf{z}^k;\phi)}\right)\right]$	вариационный вывод на основе принципа минимальной длины описания [37]

Таблица 1. Виды вариационной нижней оценки.

При максимизации по параметрам (ϕ,θ) вариационные нижние оценки из таблицы 1 будут по–разному приближать $\mathrm{q}(\mathbf{z};\phi)$ к $\mathrm{p}(\mathbf{z}|\mathbf{x};\theta)$, с разной степенью уверенности, что значительно сказывается на качестве решения конечной задачи, для которой вариационный вывод был промежуточной процедурой [34].

4 Настройка параметров в процессах диффузии

В данном разделе рассматривается динамика, заданная следующим расширенным дифференциальным уравнением:

$$\begin{cases} F([\theta^{\mathrm{T}},\mathbf{z}^{\mathrm{T}}]^{\mathrm{T}}(t),t) = \begin{bmatrix}\mathbf{0}_{d_\theta}\\ f(\mathbf{z}(t),t,\theta)\end{bmatrix}; \\ \widetilde{G}([\theta^{\mathrm{T}},\mathbf{z}^{\mathrm{T}}]^{\mathrm{T}}(t),t) = \begin{bmatrix}\mathbf{0}_{d_\theta\times m}\\ G(\mathbf{z}(t),t,\theta)\end{bmatrix}; \\ \mathrm{d}\begin{bmatrix}\theta\\ \mathbf{z}\end{bmatrix}(t) = F([\theta^{\mathrm{T}},\mathbf{z}^{\mathrm{T}}]^{\mathrm{T}}(t),t)\,\mathrm{d}\,t + \widetilde{G}([\theta^{\mathrm{T}},\mathbf{z}^{\mathrm{T}}]^{\mathrm{T}}(t),t)\,\mathrm{d}\,W_t. \end{cases}$$

$$\begin{bmatrix}\theta\\ \mathbf{z}\end{bmatrix}(t) \in \mathbb{R}^{d_\theta+d_{\mathbf{z}}} \text{ — вектор–столбец,}$$

составленный из векторов–столбцов θ и $\mathbf{z}(t)$. Для оценки градиентов по параметрам θ в процедуре градиентной оптимизации модели, заданной системой (2), рассмотрим расширенную стохастическую задачу Коши для (2):

$$\begin{cases} \mathrm{d}\begin{bmatrix}\theta\\ \mathbf{z}\end{bmatrix}(t) = F([\theta^{\mathrm{T}}, \mathbf{z}^{\mathrm{T}}]^{\mathrm{T}}(t), t)\,\mathrm{d}\,t + \widetilde{G}([\theta^{\mathrm{T}}, \mathbf{z}^{\mathrm{T}}]^{\mathrm{T}}(t), t)\,\mathrm{d}\,W_t = \begin{bmatrix}\mathbf{0}_{d_\theta}\\ f(\mathbf{z}(t), t, \theta)\end{bmatrix}\mathrm{d}\,t + \begin{bmatrix}\mathbf{0}_{d_\theta \times m}\\ G(\mathbf{z}(t), t, \theta)\end{bmatrix}\mathrm{d}\,W_t; \\ \begin{bmatrix}\theta\\ \mathbf{z}\end{bmatrix}(0) = \begin{bmatrix}\theta\\ \mathbf{z}_0\end{bmatrix}. \end{cases} \tag{17}$$

Введём обозначение: $\Theta(t) = [\theta^{\mathrm{T}}, \mathbf{z}^{\mathrm{T}}]^{\mathrm{T}}(t) \in \mathbb{R}^{d_\theta + d_{\mathbf{z}}}$, связывающее несколько переменных векторов–столбцов в один вектор–столбец, также введём

$$b(t) = \frac{\mathrm{d}\,\Theta(t)}{\mathrm{d}\,\Theta(0)}.$$

Таким образом, согласно (4):

$$\frac{\mathrm{d}\,L}{\mathrm{d}\,\Theta(0)} = \frac{\mathrm{d}\,\mathbb{E}_\mu\left[l(\mathbf{z}(T))\right]}{\mathrm{d}\,\Theta(0)}.$$

Следующее утверждение определят способ вычисления производной (4) по параметрам $\Theta(0)$.

Теорема 1. *Пусть*

$$\mathrm{d}\,\Theta(t) = F(\Theta(t), t)\,\mathrm{d}\,t + \widetilde{G}(\Theta(t), t)\,\mathrm{d}\,W_t \tag{18}$$

является многомерным процессом Ито. Функция F — дважды непрерывно дифференцируемая, $\widetilde{G}$ — дважды непрерывно дифференцируемая функция, обе функции непрерывны по Липшицу, для составляющих F и $\widetilde{G}$ функций выполнены условия регулярности (3)*. Тогда для нейросетевой модели, заданной уравнением* (18)*, необходимый для настройки параметров градиент при оптимизации* (4) *вычисляется следующим образом:*

$$\begin{cases} \dfrac{\mathrm{d}\,L}{\mathrm{d}\,\Theta(0)} = \mathbb{E}_\mu\left[\dfrac{\partial l(\mathbf{z}(T))}{\partial [\theta^{\mathrm{T}}, \mathbf{z}^{\mathrm{T}}]^{\mathrm{T}}(T)} b(T)\right]; \\ b(T) = I_{d_\theta + d_{\mathbf{z}}} + \displaystyle\int_0^T \left(\frac{\partial F(\Theta(s), s)}{\partial \Theta(s)}\,\mathrm{d}\,s + \frac{\partial(\widetilde{G}(\Theta(s), s)\,\mathrm{d}\,W_s)}{\partial \Theta(s)}\right) b(s). \end{cases} \tag{19}$$

Доказательство. Воспользовавшись теоремой Лебега о мажорируемой сходимости [38], выполним преобразование градиента математического ожидания в (4):

$$\frac{\mathrm{d}\,L}{\mathrm{d}\,\Theta(0)} = \frac{\mathrm{d}\,\mathbb{E}_\mu\left[l(\mathbf{z}(T))\right]}{\mathrm{d}\,\Theta(0)} = \mathbb{E}_\mu\left[\frac{\partial l(\mathbf{z}(T))}{\partial [\theta^{\mathrm{T}}, \mathbf{z}^{\mathrm{T}}]^{\mathrm{T}}(T)} b(T)\right].$$

Чтобы вывести выражение для $b(T)$, рассмотрим следующий процесс диффузии:

$$\mathbf{z}(T) = \mathbf{z}_0 + \int_0^T \left(f(\mathbf{z}(s), s, \theta)\,\mathrm{d}\,s + G(\mathbf{z}(s), s, \theta)\,\mathrm{d}\,W_s\right) = \widetilde{\mathbf{z}}(T, \mathbf{z}_0, \theta).$$

Применяя утверждение 2.3.1 из [39] (Proposition 2.3.1), получим $\dfrac{\mathrm{d}\,\mathbf{z}(T)}{\mathrm{d}\,\theta}$:

$$\begin{aligned}\frac{\mathrm{d}\,\widetilde{\mathbf{z}}(T,\mathbf{z}_0,\theta)}{\mathrm{d}\,\theta} &= \int\limits_0^T \frac{\mathrm{d}}{\mathrm{d}\,\theta}\left(f(\widetilde{\mathbf{z}}(s,\mathbf{z}_0,\theta),s,\theta)\,\mathrm{d}\,s + G(\widetilde{\mathbf{z}}(s,\mathbf{z}_0,\theta),s,\theta)\,\mathrm{d}\,W_s\right) = \\ &= \int\limits_0^T \left(\frac{\partial f(\widetilde{\mathbf{z}}(s,\mathbf{z}_0,\theta),s,\theta)}{\partial\theta} + \frac{\partial f(\widetilde{\mathbf{z}}(s,\mathbf{z}_0,\theta),s,\theta)}{\partial\widetilde{\mathbf{z}}}\frac{\partial\widetilde{\mathbf{z}}(s,\mathbf{z}_0,\theta)}{\partial\theta}\right)\mathrm{d}\,s + \\ &+ \int\limits_0^T \left(\frac{\partial(G(\widetilde{\mathbf{z}}(s,\mathbf{z}_0,\theta),s,\theta)\,\mathrm{d}\,W_s)}{\partial\theta} + \frac{\partial(G(\widetilde{\mathbf{z}}(s,\mathbf{z}_0,\theta),s,\theta)\,\mathrm{d}\,W_s)}{\partial\widetilde{\mathbf{z}}}\frac{\partial\widetilde{\mathbf{z}}(s,\mathbf{z}_0,\theta)}{\partial\theta}\right).\end{aligned} \tag{20}$$

Аналогично выводится $\dfrac{\mathrm{d}\,\mathbf{z}(T)}{\mathrm{d}\,\mathbf{z}_0}$:

$$\begin{aligned}\frac{\mathrm{d}\,\widetilde{\mathbf{z}}(T,\mathbf{z}_0,\theta)}{\mathrm{d}\,\mathbf{z}_0} &= \int\limits_0^T \left(\frac{\partial f(\widetilde{\mathbf{z}}(s,\mathbf{z}_0,\theta),s,\theta)}{\partial\widetilde{\mathbf{z}}}\frac{\partial\widetilde{\mathbf{z}}(s,\mathbf{z}_0,\theta)}{\partial\mathbf{z}_0}\right)\mathrm{d}\,s + \\ &+ \int\limits_0^T \left(\frac{\partial(G(\widetilde{\mathbf{z}}(s,\mathbf{z}_0,\theta),s,\theta)\,\mathrm{d}\,W_s)}{\partial\widetilde{\mathbf{z}}}\frac{\partial\widetilde{\mathbf{z}}(s,\mathbf{z}_0,\theta)}{\partial\mathbf{z}_0}\right).\end{aligned} \tag{21}$$

Используя (20), (21), восстановим вид $b(T)$:

$$b(T) = \begin{bmatrix} I_{d_\theta} & \mathbf{0}_{d_\theta\times d_\mathbf{z}} \\ \dfrac{\mathrm{d}\,\widetilde{\mathbf{z}}(T,\mathbf{z}_0,\theta)}{\mathrm{d}\,\theta} & \dfrac{\mathrm{d}\,\widetilde{\mathbf{z}}(T,\mathbf{z}_0,\theta)}{\mathrm{d}\,\mathbf{z}_0} \end{bmatrix},$$

что соответствует решению стохастической задачи Коши в момент T:

$$b(T) = I_{d_\theta+d_\mathbf{z}} + \int\limits_0^T \left(\frac{\partial F(\Theta(s),s)}{\partial\Theta(s)}\,\mathrm{d}\,s + \frac{\partial(\widetilde{G}(\Theta(s),s)\,\mathrm{d}\,W_s)}{\partial\Theta(s)}\right) b(s).$$

□

Результат теоремы 1 естественным образом обобщается на случай, когда функция потерь $l(\cdot)$ зависит явно от нескольких моментов t_i:

$$0 < t_1 < \cdots < t_i < \cdots < t_n \leq T;$$

$$L = \mathbb{E}_\mu\left[l(\mathbf{z}(t_1),\ \ldots,\ \mathbf{z}(t_n))\right];$$

$$\frac{\mathrm{d}\,L}{\mathrm{d}\,\Theta(0)} = \mathbb{E}_\mu\left[\sum_{i=1}^n \frac{\partial l(\mathbf{z}(t_1),\ \ldots,\ \mathbf{z}(t_n))}{\partial[\theta^\mathrm{T},\mathbf{z}^\mathrm{T}]^\mathrm{T}(t_i)} b(t_i)\right].$$

При выводе в условиях (17) была определена константная динамика для параметра θ. Однако для данной переменной можно ввести свою нетривиальную динамику, зависящую от своего другого параметра, как, например, было сделано в [40].

Также стоит заметить, что результаты, аналогичные теореме 1, можно получить с помощью теории оптимального управления, воспользовавшись стохастическим аналогом принципа максимума Понтрягина [22, 41].

В градиентной оптимизации (4) часто применимая

$$\text{несмещённая Монте–Карло оценка } \frac{\mathrm{d}\,\hat{L}}{\mathrm{d}\,\Theta(0)} \text{ на градиент } \frac{\mathrm{d}\,L}{\mathrm{d}\,\Theta(0)}$$

выглядит следующим образом [29]:

$$\frac{\mathrm{d}\,\hat{L}}{\mathrm{d}\,\theta} = \frac{\mathrm{d}\,l(\mathbf{z}(T))}{\mathrm{d}\,\theta} = \frac{\partial l(\mathbf{z}(T))}{\partial \mathbf{z}(T)} \frac{\mathrm{d}\,\mathbf{z}(T)}{\mathrm{d}\,\theta},\ \frac{\mathrm{d}\,\hat{L}}{\mathrm{d}\,\mathbf{z}_0} = \frac{\mathrm{d}\,l(\mathbf{z}(T))}{\mathrm{d}\,\mathbf{z}_0} = \frac{\partial l(\mathbf{z}(T))}{\partial \mathbf{z}(T)} \frac{\mathrm{d}\,\mathbf{z}(T)}{\mathrm{d}\,\mathbf{z}_0};$$

$$\frac{\mathrm{d}\,L}{\mathrm{d}\,\Theta(0)} = \mathbb{E}_\mu \left[\frac{\mathrm{d}\,\hat{L}}{\mathrm{d}\,\Theta(0)} \right] = \mathbb{E}_\mu \left[\left[\frac{\mathrm{d}\,\hat{L}}{\mathrm{d}\,\theta} \quad \frac{\mathrm{d}\,\hat{L}}{\mathrm{d}\,\mathbf{z}_0} \right] \right];$$

$\frac{\mathrm{d}\,\hat{L}}{\mathrm{d}\,\Theta(0)} \in \mathbb{R}^{d_\theta + d_\mathbf{z}}$ является вектором–строкой,

составленной из двух векторов–строк Монте–Карло оценок $\frac{\mathrm{d}\,\hat{L}}{\mathrm{d}\,\theta}$ и $\frac{\mathrm{d}\,\hat{L}}{\mathrm{d}\,\mathbf{z}_0}$.

Решение системы (19) на практике с произвольной наперёд заданной точностью возможно с помощью описанной в подразделе 3.2.2 схемы Эйлера–Маруямы, однако несмотря на теоретическую универсальность выведенной системы (19), в силу слабой масштабируемости по памяти (необходимо на каждой итерации содержать матрицу $b(t),\ t \in [0, T]$) на разностных схемах относительно небольшого размера решение системы (19) методом Эйлера–Маруямы больше требует памяти и времени вычисления одной итерации градиентного метода оптимизации, чем *методом обратного распространения ошибки через разностную схему (backprop through solver)* [12, 14].

5 Вероятностное моделирование конечных последовательностей

При построении описания временного ряда получаемая модель часто представляет собой классический пример динамической системы. Подобные модели возможно строить авторегрессионным способом, однако такие модели часто трудно поддаются интерпретации, в отличие от моделей с пространством состояний, в которых динамика между наблюдениями описывается непосредственно с помощью скрытых переменных [10]. В работе рассматриваются именно такие модели со скрытыми переменными.

Для описания подобных моделей введём несколько определений:

Определение 1. *Случайный процесс* $\mathbf{Z}(\omega),\ \omega \in \Omega$, *на* $(\Omega, \mathcal{F})$ *называется* ***прогрессивно измеримым*** *относительно* $\mathbb{F}$, *если* $\forall (s,t) \in [0,T]^2,\ 0 \leq s \leq t$ *отображение* $(s,\omega) \to \mathbf{Z}_s(\omega) : ([0,t] \times \Omega, \mathcal{B}([0,t]) \otimes \mathcal{F}_t) \to (\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$ *является измеримым.*

Определение 2. $\mathcal{A}$ *— класс* $\mathcal{F}_t$*–прогрессивно измеримых процессов*

$$\widetilde{u}(\omega) = \left\{ \widetilde{u}_t(\omega) = \left({}_1\widetilde{u}_t(\omega),\ \ldots,\ {}_m\widetilde{u}_t(\omega) \right)^{\mathrm{T}}; t \in [0,T] \right\},\ \omega \in \Omega,$$

удовлетворяющих:

$$\mathbb{E}_\mu \left[\int_0^T \left({}_i\widetilde{u}_t(\omega) \right)^2 dt \right] < \infty,\ i = \overline{1,m}.$$

$\mathcal{A}_b$ *— класс ограниченных* $\mathcal{F}_t$*–прогрессивно измеримых процессов* $\widetilde{u}$*:* $\exists R < \infty,\ \left\| \widetilde{u}_t(\omega) \right\|_2 \leq R,\ \forall (\omega, t) \in \Omega \times [0,T]$.

Определение 3. *Случайный процесс* $\mathbf{Z}(\omega),\ \omega \in \Omega$, *на* $(\Omega, \mathcal{F})$ *называется* ***простым****, если* $\exists R < \infty$, $\exists$ *строго возрастающая последовательность* $0 = t_0 < \cdots < t_j = T$, $\exists$ *последовательность случайных величин* $\xi_0,\ \ldots,\ \xi_{j-1} : \xi_i$ *является* $\mathcal{F}_{t_i}$*–измеримым,* $i = \overline{0, j-1}$, $\sup\limits_{\omega \in \Omega} \max\limits_{i \in \{0,\ \ldots,\ j-1\}} \|\xi_i(\omega)\|_2 \leq R$ *и*

$$\mathbf{Z}_t(\omega) = \xi_0(\omega) \mathbb{1}_{\{0\}}(t) + \sum_{i=0}^{j-1} \xi_i(\omega) \mathbb{1}_{(t_i, t_{i+1}]}(t),\ t \in [0,T],\ \omega \in \Omega.$$

$\mathcal{A}_s$ *— класс* ***простых*** *процессов.*

Введённые классы случайных процессов образуют цепь вложений: $\mathcal{A}_s \subset \mathcal{A}_b \subset \mathcal{A}$. В добавок к определениям введём основополагающие для текущего раздела утверждения:

Лемма 4 ([42])**.** *Пусть* $\mathbf{Z}(\omega), \omega \in \Omega$ *— ограниченный прогрессивно измеримый процесс:* $\exists R < \infty,\ \|\mathbf{Z}_t(\omega)\|_2 \leq R,\ \forall (\omega, t) \in \Omega \times [0,T]$. *Пусть* μ *— вероятностная мера, определённая на* $(\Omega, \mathcal{F})$. *Тогда существует последовательность простых процессов* ${}^i\mathbf{Z}(\omega),\ i \in \mathbb{N}$: $\sup\limits_{i \in \mathbb{N}} \left\| {}^i\mathbf{Z}_t(\omega) \right\|_2 \leq R,\ \forall (\omega, t) \in \Omega \times [0,T]$ *и*

$$\lim_{i \to \infty} \mathbb{E}_\mu \left[\int_0^T \left\| {}^i\mathbf{Z}_t(\omega) - \mathbf{Z}_t(\omega) \right\|_2^2 dt \right] = 0.$$

Лемма 5 ([43])**.** *Пусть* $(\Omega, \mathcal{F})$ *— измеримое пространство с польским пространством* Ω *и связанной с ним борелевской* σ*–алгеброй* $\mathcal{F}$. *Пусть* μ *— вероятностная мера, определённая на* $(\Omega, \mathcal{F})$ *и* $f : \Omega \to \mathbb{R}$ *— ограниченная борелевская функция,* $\mu_i,\ i \in \mathbb{N}$ *— последовательность вероятностных мер на* $(\Omega, \mathcal{F})$: $\exists R < \infty,\ \sup\limits_{i \in \mathbb{N}} \mathrm{KL}\left[\mu_i || \mu\right] \leq R$. *Предположим слабую сходимость* μ_i *к* μ. *Тогда выполнены:*

(a) $\lim\limits_{i\to\infty}\mathbb{E}_{\mu_i}\left[f(\omega)\right]=\mathbb{E}_{\mu}\left[f(\omega)\right]$;

(b) *если* $f_i,\ i\in\mathbb{N}$ *— последовательность равномерно ограниченных функций, сходящихся почти всюду к* f, *тогда* $\lim\limits_{i\to\infty}\mathbb{E}_{\mu_i}\left[f_i(\omega)\right]=\mathbb{E}_{\mu}\left[f(\omega)\right]$.

5.1 Вариационный вывод в процессах диффузии

В текущем подразделе введём для отображений $f,\ G$ отображения $\hat{f},\ \hat{G}$. Отображение, записанное вместе с символом «$\hat{}$», обладает теми же свойствами, той же сигнатурой, что и отображение, записанное без «$\hat{}$», однако такие отображения введены для того, чтобы они выполняли роль других представителей своих классов отображений. В работе вводится следующая вероятностная модель со скрытыми переменными:

$$\begin{cases}\mathbf{z}(t_0)\sim\mathrm{p}(\cdot;\theta),\ t_0=0;\\ \mathrm{d}\,\mathbf{z}(t)=\hat{f}(\mathbf{z}(t),t,\theta)\,\mathrm{d}\,t+\hat{G}(\mathbf{z}(t),t,\theta)\,\mathrm{d}\,W_t;\\ \mathbf{x}_{t_i}\sim\mathrm{p}(\cdot|\mathbf{z}(t_i);\theta),\ t_i\in[0,T],\ t_{i-1}<t_i,\ i=\overline{1,N},\ t_N=T.\end{cases}\tag{22}$$

Для модели (22) определена функция полного правдоподобия в форме элементарного объёма в терминах теории меры:

$$\mathrm{P}((\mathrm{d}\,\mathbf{x}_{t_i})_{i=1}^N,(\mathrm{d}\,\mathbf{z}(t_j))_{j=0}^N;\theta)=\prod_{i=1}^N\left(\mathrm{p}(\mathbf{x}_{t_i}|\mathbf{z}(t_i);\theta)\,\mathrm{d}\,\mathbf{x}_{t_i}\right)\mu(\mathrm{d}\,\omega)\,\mathrm{p}(\mathbf{z}_0;\theta)\,\mathrm{d}\,\mathbf{z}_0;$$

$$\int\limits_{\mathbb{R}^{d_{\mathbf{x}}N}\times\mathbb{R}^{d_{\mathbf{z}}}\times\mathbb{W}}\prod_{i=1}^N\left(\mathrm{p}(\mathbf{x}_{t_i}|\mathbf{z}(t_i);\theta)\,\mathrm{d}\,\mathbf{x}_{t_i}\right)\mathrm{p}(\mathbf{z}_0;\theta)\,\mathrm{d}\,\mathbf{z}_0\mu(\mathrm{d}\,\omega)=1,\ \omega(\cdot)\text{ — траектория реализации }W.$$

Принципиальное отличие (22) от рассмотренных моделей в [10, 20] состоит в параметризации динамики скрытых переменных с помощью стохастического дифференциального уравнения. При такой параметризации начальное скрытое состояние $\mathbf{z}(0)$ отображается через суперпозицию рекуррентных преобразований, описывая траектории в случайном поле, в котором скрытое состояние скорректированно ведёт себя относительно детерминированного преобразования, то есть преобразования с $G(\cdot)\equiv\mathbf{0}_{d_{\mathbf{z}}\times m},\ \hat{G}(\cdot)\equiv\mathbf{0}_{d_{\mathbf{z}}\times m}$, данное поведение скрытого состояния может благоприятно сказаться на качестве решения конечной задачи. Также модель в (22) отличается от рассматриваемых в [14] наличием независимого от $\mathbf{x}_{t_i}, i=\overline{1,N}$, априорного распределения. Модель (22), как и модель в [10], представляет собой рекуррентную нейронную сеть для обработки конечных последовательностей, в которой скрытое состояние между наблюдениями интерполируется с помощью динамики, заданной дифференциальным уравнением. В (22) априорное распределение задаётся следующим образом:

$$\mathrm{P}(\mathrm{d}\,\mathbf{z}_0,\mathrm{d}\,\omega;\theta)=\mathrm{p}(\mathbf{z_0};\theta)\mu(\mathrm{d}\,\omega)\,\mathrm{d}\,\mathbf{z}_0.$$

Для вывода в вероятностной модели, параметризованной процессом диффузии, необходимо определить вариационное приближение апостериорного распределения в любой рассматриваемый момент времени t. В работе такое вариационное приближение для построения процедуры вариационного вывода задаётся другим винеровским процессом:

$$\mathrm{Q}(\mathrm{d}\,\mathbf{z}_0, \mathrm{d}\,\omega|(\mathbf{x}_{t_j})_{j=1}^S; \phi) = \mathrm{q}(\mathbf{z}_0|(\mathbf{x}_{t_j})_{j=1}^S; \phi)\nu(\mathrm{d}\,\omega)\,\mathrm{d}\,\mathbf{z}_0,\ \nu \text{ — винеровская мера в } \mathbb{W},\ S \leq N.$$

Оба распределения являются корректно заданными, собственными, вероятностными:

$$\mathbb{E}_{\mathrm{P}}\left[1\right] = \mathbb{E}_{\mathrm{p}(\mathbf{z_0};\theta)}\left[\mathbb{E}_{\mu}\left[1\right]\right] = \int\limits_{\mathbb{R}^{d_{\mathbf{z}}}\times\mathbb{W}} \mathrm{P}(\mathrm{d}\,\mathbf{z}_0, \mathrm{d}\,\omega; \theta) = 1;$$

$$\mathbb{E}_{\mathrm{Q}}\left[1\right] = \mathbb{E}_{\mathrm{q}(\mathbf{z}_0|(\mathbf{x}_{t_j})_{j=1}^S;\phi)}\left[\mathbb{E}_{\nu}\left[1\right]\right] = \int\limits_{\mathbb{R}^{d_{\mathbf{z}}}\times\mathbb{W}} \mathrm{Q}(\mathrm{d}\,\mathbf{z}_0, \mathrm{d}\,\omega|(\mathbf{x}_{t_j})_{j=1}^S; \phi) = 1.$$

Таким образом, вариационная нижняя оценка на основе дивергенции α–Реньи выводится аналогично (16):

$$\begin{cases} \mathcal{L}_{\alpha}^{K}((\mathbf{x}_{t_i})_{i=1}^N; \phi, \theta) = \dfrac{1}{1-\alpha}\mathbb{E}_{\mathrm{Q}^{1:K}}\left[\ln\left(\dfrac{1}{K}\sum\limits_{k=1}^{K}\left(\dfrac{\prod\limits_{i=1}^{N}\mathrm{p}(\mathbf{x}_{t_i}|\mathbf{z}^k(t_i);\theta)\,\mathrm{P}(\mathrm{d}\,\mathbf{z}_0^k, \mathrm{d}\,\omega^k;\theta)}{\mathrm{Q}(\mathrm{d}\,\mathbf{z}_0^k, \mathrm{d}\,\omega^k|(\mathbf{x}_{t_j})_{j=1}^S;\phi)}\right)^{1-\alpha}\right)\right], \\ \mathrm{Q}^{1:K}(\mathrm{d}\,\mathbf{z}_0^{1:K}, \mathrm{d}\,\omega^{1:K}|(\mathbf{x}_{t_j})_{j=1}^S;\phi) = \prod\limits_{k=1}^{K}\mathrm{Q}(\mathrm{d}\,\mathbf{z}_0^k, \mathrm{d}\,\omega^k|(\mathbf{x}_{t_j})_{j=1}^S;\phi), \\ \mathrm{Q}^{1:K} \text{ — вероятностный закон для } K \text{ реализаций процесса диффузии.} \end{cases} \tag{23}$$

В работе рассматривается следующая параметризация процесса диффузии $\mathbf{z}^k(t)$, соответствующего ν:

$$\begin{cases} \mathrm{d}\,\mathbf{z}^k(t) = \hat{h}(\mathbf{z}^k(t), t, \phi, \widetilde{\mathbf{x}})\,\mathrm{d}\,t + \hat{G}(\mathbf{z}^k(t), t, \theta)\,\mathrm{d}\,\widetilde{W}_t^k; \\ \mathbf{z}^k(0) = \mathbf{z}_0^k \overset{\text{i.i.d.}}{\sim} \mathrm{q}(\cdot|(\mathbf{x}_{t_j})_{j=1}^S;\phi),\ k = \overline{1,K}. \end{cases} \tag{24}$$

$\widetilde{W}^k$ — k–ая реализация винеровского процесса $\widetilde{W}$, соответствующего мере ν, функция $\hat{h} : \mathbb{R}^{d_{\mathbf{z}}} \times \mathbb{R}^{+} \times \mathbb{R}^{d_\phi} \times \mathbb{R}^{d_{\widetilde{\mathbf{x}}}} \to \mathbb{R}^{d_{\mathbf{z}}}$ обладает теми же свойствами, что и $\hat{f}$; $\widetilde{\mathbf{x}}$ — переменная, агрегирующая в себе информацию о моделируемой последовательности $(\mathbf{x}_{t_i})_{i=1}^N$, например, $\widetilde{\mathbf{x}} = [\mathbf{x}_{t_1}^{\mathrm{T}}, \mathbf{x}_{t_2}^{\mathrm{T}}, \mathbf{x}_{t_3}^{\mathrm{T}}]^{\mathrm{T}}$. В дополнение к (24) предположим, что существует измеримая вспомогательная функция $\hat{u} : \mathbb{R}^{d_{\mathbf{z}}} \times \mathbb{R}^{+} \times \mathbb{R}^{d_{\widetilde{\mathbf{x}}}} \to \mathbb{R}^{m}$ такая, что:

$$\begin{cases} \hat{G}(\mathbf{z}, t, \theta)\hat{u}(\mathbf{z}, t, \widetilde{\mathbf{x}}) = \hat{h}(\mathbf{z}, t, \phi, \widetilde{\mathbf{x}}) - \hat{f}(\mathbf{z}, t, \theta),\ \forall(\mathbf{z}, t, \theta, \phi, \widetilde{\mathbf{x}}) \in \mathbb{R}^{d_{\mathbf{z}}} \times \mathbb{R}^{+} \times \mathbb{R}^{d_\theta} \times \mathbb{R}^{d_\phi} \times \mathbb{R}^{d_{\widetilde{\mathbf{x}}}}; \\ \mathbb{E}_{\mu}\left[\exp\left(\int_0^T \frac{1}{2}\left\|\hat{u}(\mathbf{z}^k(t), t, \widetilde{\mathbf{x}})\right\|_2^2 \mathrm{d}\,t\right)\right] < \infty,\ \forall \mathbf{z}_0^k \in \mathbb{R}^{d_{\mathbf{z}}},\ k = \overline{1,K}. \end{cases} \tag{25}$$

Тогда меры μ и ν связаны следующим соотношением [30] (Theorem 8.6.4), [44]:

$$\begin{cases} \nu(\mathrm{d}\,\omega^k) = M_{T,k}\mu(\mathrm{d}\,\omega^k); \\ M_{T,k} = \exp\left(\int\limits_0^T \left(-\frac{1}{2}\left\|\hat{u}(\mathbf{z}^k(t), t, \widetilde{\mathbf{x}})\right\|_2^2 \mathrm{d}\,t + \hat{u}(\mathbf{z}^k(t), t, \widetilde{\mathbf{x}})^{\mathrm{T}}\,\mathrm{d}\,W_t^k\right)\right). \end{cases}$$

Также функция $\hat{u}$ позволяет связать $\widetilde{W^k}$ с W^k (W^k — k–ая реализация винеровского процесса W, соответствующего мере μ, $\omega^k(\cdot)$ — траектория W^k в $\mathbb{W}$) с помощью выражения, называемого формулой Кэмерона–Мартина–Гирсанова [43–45]:

$$\widetilde{W_t^k} = W_t^k - \int\limits_0^t \hat{u}(\mathbf{z}^k(s), s, \widetilde{\mathbf{x}})\,\mathrm{d}\,s,\ \ t \in [0, T]. \tag{26}$$

Условия (25) позволяют определить случайный процесс $M_{T,k}$ как производную Радона–Никодима, связывающую меры μ и ν. Данный результат известен под названием теоремы Гирсанова [30], её основное следствие состоит в возможности моделирования вариационного приближения апостериорного процесса диффузии, используя априорный винеровский процесс:

$$\begin{aligned} \mathrm{d}\,\mathbf{z}^k(t) &= \hat{f}(\mathbf{z}^k(t), t, \theta)\,\mathrm{d}\,t + \hat{G}(\mathbf{z}^k(t), t, \theta)\,\mathrm{d}\,W_t^k = \{\text{соотношение (26)}\} = \\ &= \hat{f}(\mathbf{z}^k(t), t, \theta)\,\mathrm{d}\,t + \hat{G}(\mathbf{z}^k(t), t, \theta)\,\mathrm{d}\left(\widetilde{W_t^k} + \int\limits_0^t \hat{u}(\mathbf{z}^k(s), s, \widetilde{\mathbf{x}})\,\mathrm{d}\,s\right) = \\ &= \hat{f}(\mathbf{z}^k(t), t, \theta)\,\mathrm{d}\,t + \hat{G}(\mathbf{z}^k(t), t, \theta)\hat{u}(\mathbf{z}^k(t), t, \widetilde{\mathbf{x}})\,\mathrm{d}\,t + \hat{G}(\mathbf{z}^k(t), t, \theta)\,\mathrm{d}\,\widetilde{W_t^k} = \\ &= \{\text{свойства } \hat{u} \text{ из (25)}\} = \hat{h}(\mathbf{z}^k(t), t, \phi, \widetilde{\mathbf{x}})\,\mathrm{d}\,t + \hat{G}(\mathbf{z}^k(t), t, \theta)\,\mathrm{d}\,\widetilde{W_t^k}. \end{aligned}$$

Фиксируя значение $\mathbf{z}_0^k$ при выражении $\widetilde{W^k}$ через W^k, получаем:

$$\begin{aligned} \mathbf{z}^k(t) &= \mathbf{z}_0^k + \int\limits_0^t \left(\hat{f}(\mathbf{z}^k(s), s, \theta)\,\mathrm{d}\,s + \hat{G}(\mathbf{z}^k(s), s, \theta)\,\mathrm{d}\,W_s^k\right) = \\ &= \mathbf{z}_0^k + \int\limits_0^t \left(\hat{h}(\mathbf{z}^k(s), s, \phi, \widetilde{\mathbf{x}})\,\mathrm{d}\,s + \hat{G}(\mathbf{z}^k(s), s, \theta)\,\mathrm{d}\,\widetilde{W_s^k}\right). \end{aligned}$$

В терминах случайных процессов априорный процесс диффузии и вариационное приближение апостериорного процесса диффузии различаются в среднем, разделяя общую диффузию

$\hat{G}$:

$$\begin{cases} \mathbf{z}^k(t) = \mathbf{z}_0^k + \int\limits_0^t \left(\hat{f}(\mathbf{z}^k(s), s, \theta)\,\mathrm{d}\,s + \hat{G}(\mathbf{z}^k(s), s, \theta)\,\mathrm{d}\,W_s^k \right), \ k = \overline{1,K}, \\ \mathbf{z}_0^k \overset{\text{i.i.d.}}{\sim} \mathrm{p}(\cdot;\theta) \text{ — априорное распределение;} \end{cases}$$

$$\begin{cases} \mathbf{z}^k(t) = \mathbf{z}_0^k + \int\limits_0^t \left(\hat{h}(\mathbf{z}^k(s), s, \phi, \widetilde{\mathbf{x}})\,\mathrm{d}\,s + \hat{G}(\mathbf{z}^k(s), s, \theta)\,\mathrm{d}\,\widetilde{W}_s^k \right), \ k = \overline{1,K}, \\ \mathbf{z}_0^k \overset{\text{i.i.d.}}{\sim} \mathrm{q}(\cdot|(\mathbf{x}_{t_j})_{j=1}^S;\phi) \text{ — вариационное приближение.} \end{cases}$$

Выражение $M_{T,k}$ можно переписать, используя соотношение (26):

$$M_{T,k} = \exp\left(\int\limits_0^T \left(\frac{1}{2} \left\| \hat{u}(\mathbf{z}^k(t), t, \widetilde{\mathbf{x}}) \right\|_2^2 \mathrm{d}\,t + \hat{u}(\mathbf{z}^k(t), t, \widetilde{\mathbf{x}})^{\mathrm{T}}\,\mathrm{d}\,\widetilde{W}_t^k \right) \right). \tag{27}$$

С помощью (27) вариационную нижнюю оценку (23) можно представить следующим образом:

$$\begin{cases} \mathcal{L}_\alpha^K((\mathbf{x}_{t_i})_{i=1}^N;\phi,\theta) = \dfrac{1}{1-\alpha}\mathbb{E}_{\mathrm{Q}^{1:K}}\left[\ln\left(\dfrac{1}{K}\sum\limits_{k=1}^K \left(\left(\prod\limits_{i=1}^N \mathrm{p}(\mathbf{x}_{t_i}|\mathbf{z}^k(t_i);\theta) \dfrac{\mathrm{p}(\mathbf{z}_0^k;\theta)}{\mathrm{q}(\mathbf{z}_0^k|(\mathbf{x}_{t_j})_{j=1}^S;\phi)} \right)^{1-\alpha} (M_{T,k})^{\alpha-1} \right) \right) \right], \\ \widetilde{W}_t^k = W_t^k - \int\limits_0^t \hat{u}(\mathbf{z}^k(s), s, \widetilde{\mathbf{x}})\,\mathrm{d}\,s, \ t \in [0,T], \ k = \overline{1,K}, \\ M_{T,k} = \exp\left(\int\limits_0^T \left(\frac{1}{2} \left\| \hat{u}(\mathbf{z}^k(t), t, \widetilde{\mathbf{x}}) \right\|_2^2 \mathrm{d}\,t + \hat{u}(\mathbf{z}^k(t), t, \widetilde{\mathbf{x}})^{\mathrm{T}}\,\mathrm{d}\,\widetilde{W}_t^k \right) \right), \ k = \overline{1,K}, \\ \mathbf{z}^k(t) = \mathbf{z}_0^k + \int\limits_0^t \left(\hat{h}(\mathbf{z}^k(s), s, \phi, \widetilde{\mathbf{x}})\,\mathrm{d}\,s + \hat{G}(\mathbf{z}^k(s), s, \theta)\,\mathrm{d}\,\widetilde{W}_s^k \right), \ t \in [0,T], \ k = \overline{1,K}, \\ \mathbf{z}_0^k \overset{\text{i.i.d.}}{\sim} \mathrm{q}(\cdot|(\mathbf{x}_{t_j})_{j=1}^S;\phi), \ k = \overline{1,K}. \end{cases}$$

Вероятностный закон, соответствующий ν, можно эквивалентно смоделировать с помощью соответствующего μ вероятностного закона [14, 43, 44]. Данный факт в работе получен в виде теоремы 2.

Теорема 2. *Пусть*

$$\hat{\mathcal{L}}\left(\hat{u}, \left(\widetilde{W}^k \right)_{k=1}^K \right) = \frac{1}{1-\alpha} \ln\left(\frac{1}{K}\sum_{k=1}^K \left(\left(\prod_{i=1}^N \mathrm{p}(\mathbf{x}_{t_i}|\mathbf{z}^k(t_i);\theta) \frac{\mathrm{p}(\mathbf{z}_0^k;\theta)}{\mathrm{q}(\mathbf{z}_0^k|(\mathbf{x}_{t_j})_{j=1}^S;\phi)} \right)^{1-\alpha} (M_{T,k})^{\alpha-1} \right) \right)$$

является борелевской функцией со случайным процессом

$$\hat{u}(\omega^k) = \left\{ \hat{u}_t(\omega^k) = \hat{u}(\mathbf{z}^k(t), t, \widetilde{\mathbf{x}}); t \in [0,T] \right\}, \ \omega^k \in \Omega, \ k = \overline{1,K},$$

тогда:

$$\begin{cases} \mathcal{L}_\alpha^K((\mathbf{x}_{t_i})_{i=1}^N;\phi,\theta) = \mathbb{E}_{\hat{\mathrm{Q}}^{1:K}}\left[\hat{\mathcal{L}}\left(\widetilde{u},\left(W^k\right)_{k=1}^K\right)\right]; \\ \widetilde{u}(\omega^k) = \left\{\widetilde{u}_t(\omega^k) = u(\mathbf{z}^k(t),t,\widetilde{\mathbf{x}}); t\in[0,T]\right\},\ \omega^k\in\Omega,\ k=\overline{1,K}; \\ \hat{\mathrm{Q}}^{1:K}(\mathrm{d}\,\mathbf{z}_0^{1:K},\mathrm{d}\,\omega^{1:K}|(\mathbf{x}_{t_j})_{j=1}^S;\phi) = \prod\limits_{k=1}^K\left(\mathrm{q}(\mathbf{z}_0^k|(\mathbf{x}_{t_j})_{j=1}^S;\phi)\mu(\mathrm{d}\,\omega^k)\,\mathrm{d}\,\mathbf{z}_0^k\right). \end{cases}$$

Доказательство. Для начала рассмотрим последовательности ограниченных случайных процессов:

$$\begin{cases} {}^r\hat{u}(\omega^k) = \left\{{}^r\hat{u}_t(\omega^k) = \hat{u}(\mathbf{z}^k(t),t,\widetilde{\mathbf{x}})\mathbb{1}_{\left\{\left\|\hat{u}(\mathbf{z}^k(t),t,\widetilde{\mathbf{x}})\right\|_2\le r\right\}}; t\in[0,T]\right\}\in\mathcal{A}_b,\ \omega^k\in\Omega,\ r\in\mathbb{N},\ k=\overline{1,K}; \\ {}^r\widetilde{u}(\omega^k) = \left\{{}^r\widetilde{u}_t(\omega^k) = u(\mathbf{z}^k(t),t,\widetilde{\mathbf{x}})\mathbb{1}_{\left\{\left\|u(\mathbf{z}^k(t),t,\widetilde{\mathbf{x}})\right\|_2\le r\right\}}; t\in[0,T]\right\}\in\mathcal{A}_b,\ \omega^k\in\Omega,\ r\in\mathbb{N},\ k=\overline{1,K}. \end{cases}$$

Текущая задача состоит в определении $u(\cdot)$. Для этого наложим следующие ограничения на $u(\cdot)$:

$$\begin{cases} G(\mathbf{z},t,\theta)u(\mathbf{z},t,\widetilde{\mathbf{x}}) = h(\mathbf{z},t,\phi,\widetilde{\mathbf{x}}) - f(\mathbf{z},t,\theta),\ \forall(\mathbf{z},t,\theta,\phi,\widetilde{\mathbf{x}})\in\mathbb{R}^{d_\mathbf{z}}\times\mathbb{R}^+\times\mathbb{R}^{d_\theta}\times\mathbb{R}^{d_\phi}\times\mathbb{R}^{d_{\widetilde{\mathbf{x}}}}; \\ \mathbb{E}_\mu\left[\exp\left(\int_0^T\frac{1}{2}\left\|u(\mathbf{z}^k(t),t,\widetilde{\mathbf{x}})\right\|_2^2\mathrm{d}\,t\right)\right] < \infty,\ \forall\mathbf{z}_0^k\in\mathbb{R}^{d_\mathbf{z}},\ k=\overline{1,K}. \end{cases} \tag{28}$$

Отображения $(f,\ G,\ h,\ u)$ обладают аналогичными свойствами и такой же сигнатурой, что и $(\hat{f},\ \hat{G},\ \hat{h},\ \hat{u})$, однако являются другими представителями своих классов отображений. Далее возьмём и зафиксируем произвольные значения $\mathbf{z}_0^k\in\mathbb{R}^{d_\mathbf{z}},\ k=\overline{1,K}$.

Ясно, что $\lim\limits_{r\to\infty}{}^r\hat{u}(\omega^k) = \hat{u}(\omega^k)\in\mathcal{A}$, $\lim\limits_{r\to\infty}{}^r\widetilde{u}(\omega^k) = \widetilde{u}(\omega^k)\in\mathcal{A}$, $\forall\omega^k\in\Omega,\ k=\overline{1,K}$, так как выполнение второго условия из (25) и (28) приводит к выполнению ограничений для процессов класса $\mathcal{A}$; ω^k соответствует $\mathbf{z}^k(t)$. При этом первое условие из (25) и (28) может не выполняться для ${}^r\hat{u}$ и ${}^r\widetilde{u}$, однако данный факт никак не влияет на истинность доказываемой теоремы. По определению производной Радона–Никодима введём меру $\nu_{{}^r\hat{u}}$, порождённую случайным процессом ${}^r\hat{u}$:

$$\nu_{{}^r\hat{u}}(A^k) = \int\limits_{A^k}\exp\left(\int\limits_0^T\left(-\frac{1}{2}\left\|{}^r\hat{u}_t(\omega^k)\right\|_2^2\mathrm{d}\,t + {}^r\hat{u}_t(\omega^k)^\mathrm{T}\,\mathrm{d}\,W_t^k\right)\right)\mu(d\omega^k),\ A^k\in\mathcal{F},\ k=\overline{1,K}.$$

По теореме Гирсанова случайный процесс ${}^r\widetilde{W}^k(\omega^k) = \left\{{}^r\widetilde{W}_t^k(\omega^k); t\in[0,T]\right\}$,

$${}^r\widetilde{W}_t^k(\omega^k) = W_t^k(\omega^k) - \int\limits_0^t{}^r\hat{u}_s(\omega^k)\,\mathrm{d}\,s,\ t\in[0,T],\ \omega^k\in\Omega,\ r\in\mathbb{N},\ k=\overline{1,K},$$

является винеровским относительно меры $\nu_{{}^r\hat{u}}$. Определим соответствующее ${}^r\widetilde{W}^k$ отображе-

ние $\mathcal{T}_{{}^r\hat{u}}(\omega^k),\ \mathcal{T}_{{}^r\hat{u}} : \mathcal{F} \to \mathcal{F}$,

$$\mathcal{T}_{{}^r\hat{u}}(\omega^k) = \left\{ \omega^k(t) - \int\limits_0^t {}^r\hat{u}_s(\omega^k)\,\mathrm{d}\,s; t \in [0,T] \right\},\ \omega^k \in \Omega,\ r \in \mathbb{N},\ k = \overline{1,K}.$$

Тогда $\mu(A^k) = \nu_{{}^r\hat{u}}(\mathcal{T}^{-1}_{{}^r\hat{u}}(A^k)),\ \mu(\mathcal{T}_{{}^r\hat{u}}(A^k)) = \nu_{{}^r\hat{u}}(A^k),\ \forall A^k \subseteq \Omega,\ k = \overline{1,K}$.

Рассмотрим вспомогательные функции:

$$\begin{cases} \hat{\mathcal{L}}_+\left(\hat{u}, \left(\widetilde{W^k}\right)_{k=1}^K\right) = \max\left\{\hat{\mathcal{L}}\left(\hat{u}, \left(\widetilde{W^k}\right)_{k=1}^K\right), 0\right\}; \\ \hat{\mathcal{L}}_-\left(\hat{u}, \left(\widetilde{W^k}\right)_{k=1}^K\right) = \min\left\{\hat{\mathcal{L}}\left(\hat{u}, \left(\widetilde{W^k}\right)_{k=1}^K\right), 0\right\}; \\ \hat{\mathcal{L}}_+^R\left(\hat{u}, \left(\widetilde{W^k}\right)_{k=1}^K\right) = \min\left\{\hat{\mathcal{L}}_+\left(\hat{u}, \left(\widetilde{W^k}\right)_{k=1}^K\right), R\right\},\ R \in \mathbb{N}; \\ \hat{\mathcal{L}}_-^R\left(\hat{u}, \left(\widetilde{W^k}\right)_{k=1}^K\right) = \max\left\{\hat{\mathcal{L}}_-\left(\hat{u}, \left(\widetilde{W^k}\right)_{k=1}^K\right), -R\right\},\ R \in \mathbb{N}; \\ \hat{\mathcal{L}}^R\left(\hat{u}, \left(\widetilde{W^k}\right)_{k=1}^K\right) = \hat{\mathcal{L}}_+^R\left(\hat{u}, \left(\widetilde{W^k}\right)_{k=1}^K\right) + \hat{\mathcal{L}}_-^R\left(\hat{u}, \left(\widetilde{W^k}\right)_{k=1}^K\right). \end{cases}$$

Тогда

$\lim\limits_{R\to\infty} \hat{\mathcal{L}}^R\left(\hat{u}, \left(\widetilde{W^k}\right)_{k=1}^K\right) = \hat{\mathcal{L}}\left(\hat{u}, \left(\widetilde{W^k}\right)_{k=1}^K\right)$, $\hat{\mathcal{L}}^R\left(\hat{u}, \left(\widetilde{W^k}\right)_{k=1}^K\right)$ — борелевская и ограниченная.

Согласно лемме 4 для каждого ограниченного случайного процесса ${}^r\hat{u}$ существует сходящаяся к нему в сильном смысле последовательность простых случайных процессов ${}^{n,r}\hat{u},\ n \in \mathbb{N},\ R' < \infty$:

$${}^{n,r}\hat{u}(\omega^k) = \left\{ {}^{n,r}\hat{u}_t(\omega^k) = \hat{\xi}^n_{n_0}(\omega^k)\mathbb{1}_{\{0\}}(t) + \sum_{i=0}^{j_n-1} \hat{\xi}^n_{n_i}(\omega^k)\mathbb{1}_{(t_{n_i}, t_{n_{i+1}}]}(t); t \in [0,T] \right\} \in \mathcal{A}_s,\ \omega^k \in \Omega,\ r \in \mathbb{N},\ k = \overline{1,K},$$

$$0 = t_{n_0} < \cdots < t_{j_n} = T,\ \sup_{\omega\in\Omega} \max_{i\in\{0,\ \ldots,\ j_n-1\}} \left\| \hat{\xi}^n_{n_i}(\omega) \right\|_2 \le R',$$

случайные величины $\hat{\xi}^n_{n_i}$ являются $\mathcal{F}_{t_{n_i}}$–измеримыми, $i = \overline{0, j_n - 1}$. Определим новое семейство случайных величин $\xi^n_{n_i},\ i = \overline{0, j_n - 1},\ n \in \mathbb{N}$:

$$\begin{cases} \xi^n_{n_0}(\omega^k) := \hat{\xi}^n_{n_0}(\omega^k); \\ \xi^n_{n_i}(\psi^k_{n_i}) := \hat{\xi}^n_{n_i}(\omega^k),\ \psi^k_{n_i}(t) = \omega^k(t) - \sum\limits_{l=0}^{j_n-1} \hat{\xi}^n_{n_l}(\omega^k)(t_{n_{l+1}} - t_{n_l})\mathbb{1}_{(t_{n_l}, t_{n_{j_n}}]}(t),\ t \in [0, t_{n_i}],\ i = \overline{1, j_n - 1}. \end{cases} \tag{29}$$

При этом функция $\psi_{n_i}^k(\cdot)$ может быть произвольной из $\mathbb{W}$ при $t > t_{n_i}$. Соответственно, $\xi_{n_i}^n$ является $\mathcal{F}_{t_{n_i}}$–измеримой, $i = \overline{0, j_n - 1}$. То есть процесс

$$^{n,r}\widetilde{u}(\omega^k) = \left\{ {}^{n,r}\widetilde{u}_t(\omega^k) = \xi_{n_0}^n(\omega^k)\mathbb{1}_{\{0\}}(t) + \sum_{i=0}^{j_n-1} \xi_{n_i}^n(\omega^k)\mathbb{1}_{(t_{n_i}, t_{n_{i+1}}]}(t); t \in [0,T] \right\},\ \omega^k \in \Omega,\ k = \overline{1,K},$$

также простой. Благодаря сильной сходимости последовательности простых случайных процессов ${}^{n,r}\hat{u},\ n \in \mathbb{N}$, последовательность простых случайных процессов ${}^{n,r}\widetilde{u},\ n \in \mathbb{N}$, сильно сходится к ограниченному процессу из $\mathcal{A}_b$. В силу произвольности отображений из $\{u,\ f,\ G,\ h\}$ обозначим предельный процесс последовательности ${}^{n,r}\widetilde{u}$ как ${}^{r}\widetilde{u}$.

По построению ${}^{n,r}\hat{u}\left(\omega^k\right) = {}^{n,r}\widetilde{u}\left(\mathcal{T}_{n,r\hat{u}}\left(\omega^k\right)\right),\ \omega^k \in \Omega$, а соответствующий случайный процесс ${}^{n,r}\widetilde{W}^k(\omega^k) = \left\{ {}^{n,r}\widetilde{W}_t^k(\omega^k); t \in [0,T] \right\},\ n \in \mathbb{N}$:

$$^{n,r}\widetilde{W}_t^k(\omega^k) = W_t^k(\omega^k) - \int_0^t {}^{n,r}\hat{u}_s(\omega^k)\,\mathrm{d}\,s,\ t \in [0,T],\ \omega^k \in \Omega,\ r \in \mathbb{N},\ k = \overline{1,K},$$

является винеровским относительно меры $\nu_{n,r\hat{u}}$. Для $\psi^k = \mathcal{T}_{n,r\hat{u}}\left(\omega^k\right),\ (\omega^k, \psi^k) \in \Omega^2$, выполнено совпадение процессов ${}^{n,r}\hat{u}(\omega^k) = {}^{n,r}\widetilde{u}\left(\psi^k\right),\ (n,r) \in \mathbb{N}^2,\ k = \overline{1,K}$. Тогда благодаря (29) $\forall A^k \in \mathcal{F},\ k = \overline{1,K}$:

$$\begin{cases} {}^{n,r}\hat{u}(\omega^k) = {}^{n,r}\widetilde{u}\left(\mathcal{T}_{n,r\hat{u}}\left(\omega^k\right)\right) = \left\{\psi^k = \mathcal{T}_{n,r\hat{u}}\left(\omega^k\right)\right\} = {}^{n,r}\widetilde{u}\left(\psi^k\right) \in B^k,\ B^k \in \mathcal{B}(\mathrm{L}^2([0,T] : \mathbb{R}^m)); \\ \nu_{n,r\hat{u}}\left(A^k\right) = \nu_{n,r\hat{u}}\left(\left\{\psi^k : \psi^k \in A^k\right\}\right) = \left\{\psi^k = \mathcal{T}_{n,r\hat{u}}\left(\omega^k\right)\right\} = \nu_{n,r\hat{u}}\left(\left\{\omega^k : \omega^k \in \mathcal{T}_{n,r\hat{u}}^{-1}\left(A^k\right)\right\}\right) = \mu(A^k). \end{cases}$$

При этом $B^k = \left\{ {}^{n,r}\widetilde{u}\left(\psi^k\right) : \psi^k \in A^k \right\}$. Таким образом, установлено совпадение распределений

$$\left({}^{n,r}\hat{u}, \left({}^{n,r}\widetilde{W}^k \right)_{k=1}^K \right),\ (n,r) \in \mathbb{N}^2,$$

относительно меры $\nu_{n,r\hat{u}}$ и $\left({}^{n,r}\widetilde{u}, \left(W^k\right)_{k=1}^K \right),\ (n,r) \in \mathbb{N}^2$, относительно меры μ, что приводит к:

$$\mathbb{E}_{\nu_{n,r\hat{u}}^{1:K}}\left[\hat{\mathcal{L}}^R\left({}^{n,r}\hat{u}, \left({}^{n,r}\widetilde{W}^k \right)_{k=1}^K \right)\right] = \mathbb{E}_{\mu^{1:K}}\left[\hat{\mathcal{L}}^R\left({}^{n,r}\widetilde{u}, \left(W^k\right)_{k=1}^K \right)\right],\ (n,r,R) \in \mathbb{N}^3. \tag{30}$$

$\mu^{1:K}$ — усреднение по K независимым реализациям случайного процесса относительно меры μ, $\nu_{n,r\hat{u}}^{1:K}$ — усреднение по K независимым реализациям случайного процесса относительно меры $\nu_{n,r\hat{u}}$.

Из сильной сходимости $\left({}^{n,r}\widetilde{u}, {}^{n,r}\hat{u} \right)$ к $\left({}^{r}\widetilde{u}, {}^{r}\hat{u} \right)$ следует сходимость по распределению

$$\left(\left({}^{n,r}\widetilde{u}, \left(W^k\right)_{k=1}^K \right), \left({}^{n,r}\hat{u}, \left({}^{n,r}\widetilde{W}^k \right)_{k=1}^K \right) \right) \text{ к } \left(\left({}^{r}\widetilde{u}, \left(W^k\right)_{k=1}^K \right), \left({}^{r}\hat{u}, \left({}^{r}\widetilde{W}^k \right)_{k=1}^K \right) \right)$$

соответственно, что позволяет применить лемму 5. Воспользовавшись леммой 5, получаем:

$$\begin{cases} \lim\limits_{n\to\infty} \mathbb{E}_{\mu^{1:K}} \left[\hat{\mathcal{L}}^R \left({}^{n,r}\widetilde{u}, \left(W^k\right)_{k=1}^K \right) \right] = \mathbb{E}_{\mu^{1:K}} \left[\hat{\mathcal{L}}^R \left({}^{r}\widetilde{u}, \left(W^k\right)_{k=1}^K \right) \right], \ (r,R) \in \mathbb{N}^2; \\ \lim\limits_{n\to\infty} \mathbb{E}_{\nu^{1:K}_{n,r\hat{u}}} \left[\hat{\mathcal{L}}^R \left({}^{n,r}\hat{u}, \left({}^{n,r}\widetilde{W}^k \right)_{k=1}^K \right) \right] = \mathbb{E}_{\nu^{1:K}_{r\hat{u}}} \left[\hat{\mathcal{L}}^R \left({}^{r}\hat{u}, \left({}^{r}\widetilde{W}^k \right)_{k=1}^K \right) \right], \ (r,R) \in \mathbb{N}^2. \end{cases} \tag{31}$$

При увеличении r процессы $\left({}^{r}\widetilde{u}, {}^{r}\hat{u}\right)$ будут приближаться к $\left(\widetilde{u}, \hat{u}\right)$ соответственно, что в условиях ограничения значений функционала $\hat{\mathcal{L}}^R(\cdot)$ отрезком $[-R, R]$ позволяет применить теорему Лебега о мажорируемой сходимости вместе с леммой 5 для окончательного доопределения $u(\cdot)$ и утверждения следующего:

$$\begin{cases} \lim\limits_{r\to\infty} \mathbb{E}_{\mu^{1:K}} \left[\hat{\mathcal{L}}^R \left({}^{r}\widetilde{u}, \left(W^k\right)_{k=1}^K \right) \right] = \mathbb{E}_{\mu^{1:K}} \left[\hat{\mathcal{L}}^R \left(\widetilde{u}, \left(W^k\right)_{k=1}^K \right) \right], \ R \in \mathbb{N}; \\ \lim\limits_{r\to\infty} \mathbb{E}_{\nu^{1:K}_{r\hat{u}}} \left[\hat{\mathcal{L}}^R \left({}^{r}\hat{u}, \left({}^{r}\widetilde{W}^k \right)_{k=1}^K \right) \right] = \mathbb{E}_{\nu^{1:K}_{\hat{u}}} \left[\hat{\mathcal{L}}^R \left(\hat{u}, \left(\widetilde{W}^k \right)_{k=1}^K \right) \right], \ R \in \mathbb{N}. \end{cases} \tag{32}$$

Для функции $\hat{\mathcal{L}}^R(\cdot)$ в силу построения применима теорема Лебега о мажорируемой сходимости, которая устанавливает следующее представление:

$$\begin{cases} \lim\limits_{R\to\infty} \mathbb{E}_{\mu^{1:K}} \left[\hat{\mathcal{L}}^R \left(\widetilde{u}, \left(W^k\right)_{k=1}^K \right) \right] = \mathbb{E}_{\mu^{1:K}} \left[\hat{\mathcal{L}} \left(\widetilde{u}, \left(W^k\right)_{k=1}^K \right) \right]; \\ \lim\limits_{R\to\infty} \mathbb{E}_{\nu^{1:K}_{\hat{u}}} \left[\hat{\mathcal{L}}^R \left(\hat{u}, \left(\widetilde{W}^k \right)_{k=1}^K \right) \right] = \mathbb{E}_{\nu^{1:K}_{\hat{u}}} \left[\hat{\mathcal{L}} \left(\hat{u}, \left(\widetilde{W}^k \right)_{k=1}^K \right) \right] = \mathbb{E}_{\nu^{1:K}} \left[\hat{\mathcal{L}} \left(\hat{u}, \left(\widetilde{W}^k \right)_{k=1}^K \right) \right]. \end{cases} \tag{33}$$

В силу произвольности $\mathbf{z}_0^k \in \mathbb{R}^{d_{\mathbf{z}}}$, $k = \overline{1,K}$, выражения выше верны для любых $\mathbf{z}_0^k \in \mathbb{R}^{d_{\mathbf{z}}}$, $k = \overline{1,K}$, что позволяет по свойствам предельного значения равенства и благодаря цепочке равенств (30), (31), (32), (33) вывести эквивалентную форму вариационной нижней оценки:

$$\begin{cases} \mathcal{L}_\alpha^K((\mathbf{x}_{t_i})_{i=1}^N; \phi, \theta) = \mathbb{E}_{\mathrm{Q}^{1:K}} \left[\hat{\mathcal{L}} \left(\hat{u}, \left(\widetilde{W}^k \right)_{k=1}^K \right) \right] = \mathbb{E}_{\hat{\mathrm{Q}}^{1:K}} \left[\hat{\mathcal{L}} \left(\widetilde{u}, \left(W^k\right)_{k=1}^K \right) \right]; \\ \hat{\mathrm{Q}}^{1:K}(\mathrm{d}\,\mathbf{z}_0^{1:K}, \mathrm{d}\,\omega^{1:K} | (\mathbf{x}_{t_j})_{j=1}^S; \phi) = \prod\limits_{k=1}^K \left(\mathrm{q}(\mathbf{z}_0^k | (\mathbf{x}_{t_j})_{j=1}^S; \phi) \mu(\mathrm{d}\,\omega^k)\, \mathrm{d}\,\mathbf{z}_0^k \right). \end{cases}$$

□

Результаты теоремы 2 позволяют оценку $\mathcal{L}_\alpha^K((\mathbf{x}_{t_i})_{i=1}^N; \phi, \theta)$ привести к эквивалентному

виду:

$$\begin{cases} \mathcal{L}_\alpha^K((\mathbf{x}_{t_i})_{i=1}^N;\phi,\theta) = \dfrac{1}{1-\alpha}\mathbb{E}_{\hat{\mathrm{Q}}^{1:K}}\left[\ln\left(\dfrac{1}{K}\sum\limits_{k=1}^K\left(\left(\prod\limits_{i=1}^N \mathrm{p}(\mathbf{x}_{t_i}|\mathbf{z}^k(t_i);\theta)\dfrac{\mathrm{p}(\mathbf{z}_0^k;\theta)}{\mathrm{q}(\mathbf{z}_0^k|(\mathbf{x}_{t_j})_{j=1}^S;\phi)}\right)^{1-\alpha}(M_{T,k})^{\alpha-1}\right)\right)\right], \\ \hat{\mathrm{Q}}^{1:K}(\mathrm{d}\,\mathbf{z}_0^{1:K},\mathrm{d}\,\omega^{1:K}|(\mathbf{x}_{t_j})_{j=1}^S;\phi) = \prod\limits_{k=1}^K\left(\mathrm{q}(\mathbf{z}_0^k|(\mathbf{x}_{t_j})_{j=1}^S;\phi)\mu(\mathrm{d}\,\omega^k)\,\mathrm{d}\,\mathbf{z}_0^k\right), \\ M_{T,k} = \exp\left(\int\limits_0^T\left(\dfrac{1}{2}\left\|u(\mathbf{z}^k(t),t,\widetilde{\mathbf{x}})\right\|_2^2\mathrm{d}\,t + u(\mathbf{z}^k(t),t,\widetilde{\mathbf{x}})^{\mathrm{T}}\,\mathrm{d}\,W_t^k\right)\right),\ k=\overline{1,K}, \\ \mathbf{z}^k(t) = \mathbf{z}_0^k + \int\limits_0^t\left(h(\mathbf{z}^k(s),s,\phi,\widetilde{\mathbf{x}})\,\mathrm{d}\,s + G(\mathbf{z}^k(s),s,\theta)\,\mathrm{d}\,W_s^k\right),\ t\in[0,T],\ k=\overline{1,K}, \\ \mathbf{z}_0^k \overset{\text{i.i.d.}}{\sim} \mathrm{q}(\cdot|(\mathbf{x}_{t_j})_{j=1}^S;\phi),\ k=\overline{1,K}. \end{cases} \tag{34}$$

Максимизация оценки $\mathcal{L}_\alpha^K((\mathbf{x}_{t_i})_{i=1}^N;\phi,\theta), \alpha \geq 0$, в (34) в общем случае из–за u представляет собой задачу условной оптимизации, однако в данной работе использование функции G с $m = d_{\mathbf{z}}$ и с ненулевыми значениями только на диагонали позволяет явно выразить функцию u, сводя задачу условной максимизации оценки (34) к задаче безусловной оптимизации:

$$u(\mathbf{z},t,\widetilde{\mathbf{x}}) = \left(\frac{h_l(\mathbf{z},t,\phi,\widetilde{\mathbf{x}}) - f_l(\mathbf{z},t,\theta)}{G_{ll}(\mathbf{z},t,\theta)}\right)_{l=1}^{d_{\mathbf{z}}},\ \forall(\mathbf{z},t,\theta,\phi,\widetilde{\mathbf{x}}) \in \mathbb{R}^{d_{\mathbf{z}}}\times\mathbb{R}^+\times\mathbb{R}^{d_\theta}\times\mathbb{R}^{d_\phi}\times\mathbb{R}^{d_{\widetilde{\mathbf{x}}}}.$$

Так же в более общем случае, с невырожденной матрицей $G(\mathbf{z},t,\theta)^{\mathrm{T}}G(\mathbf{z},t,\theta)$, $\forall(\mathbf{z},t,\theta) \in \mathbb{R}^{d_{\mathbf{z}}}\times\mathbb{R}^+\times\mathbb{R}^{d_\theta}$, функция u имеет явное представление:

$$u(\mathbf{z},t,\widetilde{\mathbf{x}}) = \left(G(\mathbf{z},t,\theta)^{\mathrm{T}}G(\mathbf{z},t,\theta)\right)^{-1}G(\mathbf{z},t,\theta)^{\mathrm{T}}\left(h(\mathbf{z},t,\phi,\widetilde{\mathbf{x}}) - f(\mathbf{z},t,\theta)\right),$$

$$\forall(\mathbf{z},t,\theta,\phi,\widetilde{\mathbf{x}}) \in \mathbb{R}^{d_{\mathbf{z}}}\times\mathbb{R}^+\times\mathbb{R}^{d_\theta}\times\mathbb{R}^{d_\phi}\times\mathbb{R}^{d_{\widetilde{\mathbf{x}}}}.$$

Аналогичная таблице 1 таблица $\mathcal{L}_\alpha^K((\mathbf{x}_{t_i})_{i=1}^N;\phi,\theta)$ для специальных значений α приведена ниже:

Значение α	**Нижняя оценка** $\ln \mathrm{p}((\mathbf{x}_{t_i})_{i=1}^N; \theta)$
0	$\mathbb{E}_{\hat{\mathrm{Q}}^{1:K}} \left[\ln \left(\frac{1}{K} \sum_{k=1}^{K} \left(\prod_{i=1}^{N} \mathrm{p}(\mathbf{x}_{t_i} \vert \mathbf{z}^k(t_i); \theta) \left(\frac{\mathrm{p}(\mathbf{z}_0^k; \theta)}{\mathrm{q}(\mathbf{z}_0^k \vert (\mathbf{x}_{t_j})_{j=1}^S; \phi) M_{T,k}} \right) \right) \right) \right]$
1	$\mathbb{E}_{\hat{\mathrm{Q}}^{1:K}} \left[\frac{1}{K} \sum_{k=1}^{K} \left(\sum_{i=1}^{N} \ln \left(\mathrm{p}(\mathbf{x}_{t_i} \vert \mathbf{z}^k(t_i); \theta) \right) - \int_0^T \frac{1}{2} \left\| u(\mathbf{z}^k(t), t, \widetilde{\mathbf{x}}) \right\|_2^2 \mathrm{d}\, t \right) \right] - \mathrm{KL} \left[\mathrm{q}(\mathbf{z}_0 \vert (\mathbf{x}_{t_j})_{j=1}^S; \phi) \vert\vert \mathrm{p}(\mathbf{z}_0; \theta) \right]$
$+\infty$	$\mathbb{E}_{\hat{\mathrm{Q}}^{1:K}} \left[\min_{k \in \{1,\ \ldots,\ K\}} \left(\sum_{i=1}^{N} \ln \left(\mathrm{p}(\mathbf{x}_{t_i} \vert \mathbf{z}^k(t_i); \theta) \right) + \left(\ln \left(\mathrm{p}(\mathbf{z}_0^k; \theta) \right) - \ln \left(\mathrm{q}(\mathbf{z}_0^k \vert (\mathbf{x}_{t_j})_{j=1}^S; \phi) \right) - \ln \left(M_{T,k} \right) \right) \right) \right]$

Таблица 2. Виды вариационной нижней оценки.

5.2 Вариационный вывод в конечных последовательностях

5.2.1 Вычисление Монте–Карло оценки вариационной нижней оценки

В процедуре вариационного вывода для модели (22) параметризация распределения

$$\mathrm{q}(\mathbf{z}_0^k \vert (\mathbf{x}_{t_j})_{j=1}^S; \phi),\ k = \overline{1, K},$$

осуществлена с помощью прохода рекуррентной нейронной сетью с архитектурой GRU [46] по сэмплам временного ряда $(\mathbf{x}_{t_j})_{j=1}^S$ с модификацией скрытого состояния, реализованной с помощью решения задачи Коши аналогично тому, как данную процедуру реализовали в [10]. При этом проход всегда выполняется, начиная с $\mathbf{x}_{t_S}$ и заканчивая $\mathbf{x}_{t_1}$. В конце последовательности нейронная сеть возвращает параметры распределения, из которого сэмплируется $\mathbf{z}_0^k$. Схематично процедура вывода скрытого представления последовательности описана ниже.

Algorithm 2 Корректирующее преобразование состояния $\mathbf{h}$ с помощью GRUCell.

function GRUCell($\mathbf{h}$, $\mathbf{x}$)

gate $= \sigma(f_{gate}([\mathbf{h}^{\mathrm{T}},\ \mathbf{x}^{\mathrm{T}}]^{\mathrm{T}}))$ ▷ поэлементное преобразование $\sigma(y) = \frac{1}{1+\exp(-y)}$

reset $= \sigma(f_{reset}([\mathbf{h}^{\mathrm{T}},\ \mathbf{x}^{\mathrm{T}}]^{\mathrm{T}}))$ ▷ $f_{gate},\ f_{reset}$ — полносвязные нейронные сети

$\mathbf{h}' = \mathrm{NN}([(\mathrm{reset} \odot \mathbf{h})^{\mathrm{T}},\ \mathbf{x}^{\mathrm{T}}]^{\mathrm{T}})$ ▷ NN — полносвязная нейронная сеть

$\mathbf{h} := (1 - \mathrm{gate}) \odot \mathbf{h}' + \mathrm{gate} \odot \mathbf{h}$

return h

Algorithm 3 Кодирование входной последовательности.

function SEQUENCEENCODE($(\mathbf{x}_{t_j})_{j=1}^S$, $(t_j)_{j=1}^S$)

$[\mathbf{h}_{t_S}^{\text{mean}^\text{T}}, \mathbf{h}_{t_S}^{\text{scale}^\text{T}}]^\text{T} = [\mathbf{0}_{d_\mathbf{z}}^\text{T}, \mathbf{1}_{d_\mathbf{z}}^\text{T}]^\text{T}$ ▷ вектор–столбец, составленный из векторов–столбцов

for $j = S,\ S-1,\ \ldots,\ 1$ **do**

$[\mathbf{h}_{t_j}^{\text{mean}^\text{T}}, \mathbf{h}_{t_j}^{\text{scale}^\text{T}}]^\text{T} := \text{GRUCell}([\mathbf{h}_{t_j}^{\text{mean}^\text{T}}, \mathbf{h}_{t_j}^{\text{scale}^\text{T}}]^\text{T}, \mathbf{x}_{t_j})$

$\ldots,\ [\mathbf{h}_{t_{j-1}}^{\text{mean}^\text{T}}, \mathbf{h}_{t_{j-1}}^{\text{scale}^\text{T}}]^\text{T} = \text{SDESolve}((\hat{\tilde{f}}, \phi), (\mathbf{0}_{2\cdot d_\mathbf{z}\times m}, \mathbf{0}_{d_\phi}), [\mathbf{h}_{t_j}^{\text{mean}^\text{T}}, \mathbf{h}_{t_j}^{\text{scale}^\text{T}}]^\text{T}, (t_j,\ \ldots,\ t_{j-1}))$

▷

$\mathrm{d}[\mathbf{h}^{\text{mean}^\text{T}}, \mathbf{h}^{\text{scale}^\text{T}}]^\text{T}(t) = \tilde{f}([\mathbf{h}^{\text{mean}^\text{T}}, \mathbf{h}^{\text{scale}^\text{T}}]^\text{T}(t), t, \phi, \tilde{\mathbf{x}})\,\mathrm{d}\,t,\ t_0 = 0,$

$\tilde{f}$ — полносвязная нейронная сеть,

$\hat{\tilde{f}}([\mathbf{h}^{\text{mean}^\text{T}}, \mathbf{h}^{\text{scale}^\text{T}}]^\text{T}(t), t, \phi) = \tilde{f}([\mathbf{h}^{\text{mean}^\text{T}}, \mathbf{h}^{\text{scale}^\text{T}}]^\text{T}(t), t, \phi, \tilde{\mathbf{x}})$

return $\mathbf{h}_{t_0}^{\text{mean}} + |\mathbf{h}_{t_0}^{\text{scale}}|\xi,\ \xi \sim \mathcal{N}(\mathbf{0}_{d_\mathbf{z}}, I_{d_\mathbf{z}})$ ▷

$\mathbf{z}_0^k \sim \mathrm{q}(\cdot|(\mathbf{x}_{t_j})_{j=1}^S; \phi);$

$|\cdot|$ — поэлементное взятие модуля;

$\mathrm{q}(\cdot|(\mathbf{x}_{t_j})_{j=1}^S; \phi) = \mathcal{N}(\mathbf{h}_{t_0}^{\text{mean}}, \text{diag}(\mathbf{h}_{t_0}^{\text{scale}} \odot \mathbf{h}_{t_0}^{\text{scale}})).$

После вычисления $\mathbf{z}_0^k$ в процедуре SequenceEncode происходит моделирование скрытых состояний $\{(\mathbf{z}^k(t_i))_{i=1}^N\}_{k=1}^K$ с помощью метода Эйлера–Маруямы, описанного в листинге 1:

$$\begin{aligned}
&1.\ \tilde{\mathbf{x}} = \text{NN'}([\mathbf{x}_{t_1}^\text{T}, \mathbf{x}_{t_2}^\text{T}, \mathbf{x}_{t_3}^\text{T}]^\text{T}),\ \{\text{NN'}, h, G\} \text{ — полносвязные нейронные сети;} \\
&2.\ \mathbf{z}_0^k \overset{\text{i.i.d.}}{\sim} \mathrm{q}(\cdot|(\mathbf{x}_{t_j})_{j=1}^S; \phi),\ g_1(\mathbf{z}^k(t), t, \phi) = h(\mathbf{z}^k(t), t, \phi, \tilde{\mathbf{x}}),\ k = \overline{1, K}; \qquad (35) \\
&3.\ \mathbf{z}^k(0),\ \mathbf{z}^k(t_1),\ \ldots,\ \mathbf{z}^k(t_N) = \text{SDESolve}((g_1, \phi), (G, \theta), \mathbf{z}_0^k, (t_0,\ t_1,\ \ldots,\ t_N)).
\end{aligned}$$

Параллельно с процедурой (35) с помощью метода Эйлера–Маруямы происходит вычисление $M_{T,k}$ с

$$u(\mathbf{z}^k(t), t, \tilde{\mathbf{x}}) = \left(\frac{h_l(\mathbf{z}^k(t), t, \phi, \tilde{\mathbf{x}}) - f_l(\mathbf{z}^k(t), t, \theta)}{G_{ll}(\mathbf{z}^k(t), t, \theta)}\right)_{l=1}^{d_\mathbf{z}}:$$

$$
\begin{aligned}
&1.\ g_2(\mathbf{z}^k(t), t, [\phi^{\mathrm{T}}, \theta^{\mathrm{T}}]^{\mathrm{T}}) = \frac{1}{2}\left\| u(\mathbf{z}^k(t), t, \widetilde{\mathbf{x}}) \right\|_2^2,\ g_3(\mathbf{z}^k(t), t, [\phi^{\mathrm{T}}, \theta^{\mathrm{T}}]^{\mathrm{T}}) = u(\mathbf{z}^k(t), t, \widetilde{\mathbf{x}})^{\mathrm{T}},\ \mathrm{I}_0^k = 0,\ k = \overline{1, K};\\
&2.\ \mathrm{I}_0^k,\ \mathrm{I}_{t_1}^k,\ \ldots,\ \mathrm{I}_{t_N}^k = \mathrm{SDESolve}((g_2, [\phi^{\mathrm{T}}, \theta^{\mathrm{T}}]^{\mathrm{T}}), (g_3, [\phi^{\mathrm{T}}, \theta^{\mathrm{T}}]^{\mathrm{T}}), \mathrm{I}_0^k, (t_0,\ t_1,\ \ldots,\ t_N));\\
&3.\ M_{T,k} = \exp(\mathrm{I}_{t_N}^k).
\end{aligned}
\tag{36}
$$

Таким образом, в результате выполнения процедур (35) и (36) вычисляются все необходимые значения для подсчёта вариационной нижней оценки на основе дивергенции α–Реньи (34) при настройке параметров модели (22) с помощью процедуры вариационного вывода. Введённые величины позволяют определить Монте–Карло оценку $\hat{\mathcal{L}}_\alpha^K((\mathbf{x}_{t_i})_{i=1}^N; \phi, \theta)$ вариационной нижней оценки $\mathcal{L}_\alpha^K((\mathbf{x}_{t_i})_{i=1}^N; \phi, \theta)$:

$$
\begin{aligned}
&\hat{\mathcal{L}}_\alpha^K((\mathbf{x}_{t_i})_{i=1}^N; \phi, \theta) = \frac{1}{1-\alpha} \ln \left(\frac{1}{K} \sum_{k=1}^K \left(\left(\prod_{i=1}^N \mathrm{p}(\mathbf{x}_{t_i} | \mathbf{z}^k(t_i); \theta) \frac{\mathrm{p}(\mathbf{z}_0^k; \theta)}{\mathrm{q}(\mathbf{z}_0^k | (\mathbf{x}_{t_j})_{j=1}^S; \phi)} \right)^{1-\alpha} (M_{T,k})^{\alpha-1} \right) \right),\\
&\mathcal{L}_\alpha^K((\mathbf{x}_{t_i})_{i=1}^N; \phi, \theta) = \mathbb{E}_{\hat{\mathrm{Q}}^{1:K}} \left[\hat{\mathcal{L}}_\alpha^K((\mathbf{x}_{t_i})_{i=1}^N; \phi, \theta) \right].
\end{aligned}
\tag{37}
$$

5.2.2 Процедура обучения модели процесса диффузии

С помощью введённых отображений, в работе определена процедура обучения модели процесса диффузии (22). Состоит эта процедура в итеративной настройке параметров градиентным методом стохастической оптимизации. На каждой итерации на разностной схеме методом Эйлера–Маруямы вычисляется Монте–Карло оценка вариационной нижней оценки (37); параметры модели обновляются шагом вдоль Монте–Карло оценки градиента вариационной нижней оценки, вычисленного с помощью обратного распространения ошибки через разностную схему [12]. Алгоритмически процедура настройки модели (22) представлена в листинге 4.

Algorithm 4 Процедура настройки модели (22).

procedure FitModel($X^{N_D}, N^e, N^b, K, \alpha, (\phi, \theta)$) ▷

N^e — количество эпох, N^b — количество батчей, N_j^{batch} — размер батча

for $1,\ 2,\ \ldots,\ N^e$ **do**

for $\mathrm{b}_j \in \mathrm{B},\ j = \overline{1, N^b}$ **do** ▷ $\mathrm{B} = \sqcup_{j=1}^{N^b}\{({}^{i_{j_l}}\mathbf{x}_{t_1},\ \ldots,\ {}^{i_{j_l}}\mathbf{x}_{t_{N_{i_{j_l}}}})\}_{l=1}^{N_j^{batch}} = X^{N_D}$

Батч конечных последовательностей $\mathrm{b}_j = \{({}^{i_{j_l}}\mathbf{x}_{t_1},\ \ldots,\ {}^{i_{j_l}}\mathbf{x}_{t_{N_{i_{j_l}}}})\}_{l=1}^{N_j^{batch}}$

Вычислить Монте–Карло оценку $\frac{1}{N_j^{batch}} \sum\limits_{l=1}^{N_j^{batch}} \hat{\mathcal{L}}_\alpha^K(({}^{i_{j_l}}\mathbf{x}_{t_o})_{o=1}^{N_{i_{j_l}}}; \phi, \theta)$

Вычислить Монте–Карло оценку $\frac{1}{N_j^{batch}} \sum\limits_{l=1}^{N_j^{batch}} \nabla_{(\phi,\theta)}\hat{\mathcal{L}}_\alpha^K(({}^{i_{j_l}}\mathbf{x}_{t_o})_{o=1}^{N_{i_{j_l}}}; \phi, \theta)$

Обновить (ϕ, θ) шагом вдоль оценки $\frac{1}{N_j^{batch}} \sum\limits_{l=1}^{N_j^{batch}} \nabla_{(\phi,\theta)}\hat{\mathcal{L}}_\alpha^K(({}^{i_{j_l}}\mathbf{x}_{t_o})_{o=1}^{N_{i_{j_l}}}; \phi, \theta)$

5.2.3 Генерация последовательности в модели процесса диффузии

Для настроенной модели (22) сэмплирование новых последовательностей из вариационного приближения и из априорного распределения состоит в следующих двух процедурах.

Процедура генерации последовательностей из аппроксимации апостериорного распределения:

1. $\mathbf{z}_0^k \overset{\text{i.i.d.}}{\sim} \mathrm{q}(\cdot|(\mathbf{x}_{t_j})_{j=1}^S; \phi)$;
2. $\mathbf{z}^k(0),\ \mathbf{z}^k(t_1),\ \ldots,\ \mathbf{z}^k(t_N) = \mathrm{SDESolve}((g_1, \phi), (G, \theta), \mathbf{z}_0^k, (0,\ t_1,\ \ldots,\ t_N))$;
3. $\mathbf{x}_{t_i}^k \sim \mathrm{p}(\cdot|\mathbf{z}^k(t_i); \theta),\ i = \overline{1, N},\ k = \overline{1, K},$ p — нормальное.

Процедура генерации последовательностей из априорного распределения:

1. $\mathbf{z}_0^k \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{0}_{d_\mathbf{z}}, I_{d_\mathbf{z}}),\ \{\hat{f}, \hat{G}\}$ — полносвязные нейронные сети;
2. $\mathbf{z}^k(0),\ \mathbf{z}^k(t_1),\ \ldots,\ \mathbf{z}^k(t_N) = \mathrm{SDESolve}((\hat{f}, \theta), (\hat{G}, \theta), \mathbf{z}_0^k, (0,\ t_1,\ \ldots,\ t_N))$;
3. $\mathbf{x}_{t_i}^k \sim \mathrm{p}(\cdot|\mathbf{z}^k(t_i); \theta),\ i = \overline{1, N},\ k = \overline{1, K},$ p — нормальное.

Схематическое описание вывода последовательности в модели (22) обобщено на рис. 1. На данном рисунке аббревиатурой ОДУ обозначена интерполяция скрытого состояния $[\mathbf{h}^{\mathrm{mean}^\mathrm{T}}, \mathbf{h}^{\mathrm{scale}^\mathrm{T}}]^\mathrm{T}(t)$ с помощью обыкновенного дифференциального уравнения, использованного в процедуре SequenceEncode:

$$\mathrm{d}[\mathbf{h}^{\mathrm{mean}^\mathrm{T}}, \mathbf{h}^{\mathrm{scale}^\mathrm{T}}]^\mathrm{T}(t) = \widetilde{f}([\mathbf{h}^{\mathrm{mean}^\mathrm{T}}, \mathbf{h}^{\mathrm{scale}^\mathrm{T}}]^\mathrm{T}(t), t, \phi, \widetilde{\mathbf{x}})\,\mathrm{d}\,t.$$

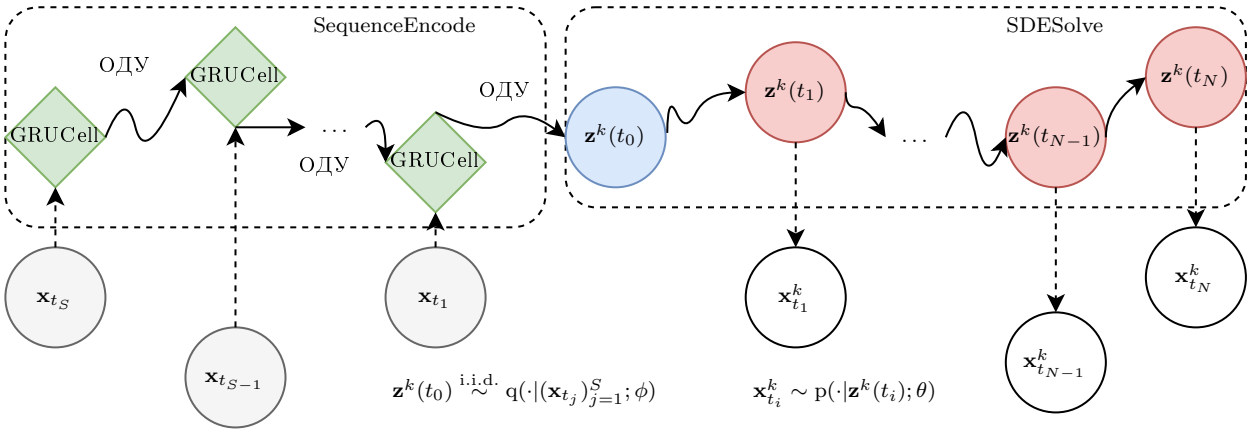

Рис. 1. Схема моделирования последовательности в модели процесса диффузии (22).

6 Вычислительные эксперименты

Для качественной оценки предложенной в данной работе модели (22) была проведена серия экспериментов на данных CMU Graphics Lab Motion Capture Database для решения задачи моделирования траектории движения человека при известной начальной траектории [47].

Данные представляют собой набор из 5 выборок, каждая из которых является набором многомерных временных рядов, относящихся к отдельному субъекту. Каждый многомерный ряд описывает движения субъекта в трёхмерном пространстве. Роль субъекта исполняет человек, снаряжённый датчиками, который выполняет относительно несложную последовательность действий, то есть каждое действие регистрируется в виде схематического скелета. Каждый временной ряд соответствует отдельному действию вида пройти из одного конца пространства в другой, сделать кувырок, присесть, встать, и тому подобное.

Для получения робастных результатов было настроено несколько моделей процесса диффузии (22), по одной на каждую выборку. Выборки отличаются размером по количеству временных рядов и по длине временного ряда. Каждая выборка также отличается видом движений человека: в одной выборке могут быть собраны различные виды приседаний, в другой выборке человек может ходить по кругу, в третьей — махать руками. Все линейные показания в выборках переведены из дюймов в метры, в целях повышения вычислительной устойчивости введена искусственная временная шкала с начальным моментом времени t_0 в нуле и конечным t_N — в единице. Краткая характеристика использованных выборок представлена в таблице 3.

Субъект № (выборка №)	Размер выборки	Количество моментов времени	Размерность ряда
83	68	99	93
79	96	130	93
138	55	316	93
69	75	343	93
80	73	660	93

Таблица 3. Характеристики использованных данных.

В экспериментах архитектуры нейронных сетей, соответствующих

$$f(\mathbf{z}^k(t), t, \theta),\ G(\mathbf{z}^k(t), t, \theta),\ h(\mathbf{z}^k(t), t, \phi, \widetilde{\mathbf{x}}),\ \mathrm{p}(\mathbf{x}_{t_i}|\mathbf{z}^k(t_i); \theta),$$

взяты из работы [14], архитектура нейронной сети, параметризующей $\mathrm{q}(\mathbf{z}_0^k|(\mathbf{x}_{t_j})_{j=1}^S; \phi)$, взята из работы [10]. Архитектура использованных в работе отображений представлена на рис. 2 и 3. Распределения $\mathrm{p}(\mathbf{z}_0^k; \theta),\ \mathrm{p}(\mathbf{x}_{t_i}|\mathbf{z}^k(t_i); \theta),\ \mathrm{q}(\mathbf{z}_0^k|(\mathbf{x}_{t_j})_{j=1}^S; \phi)$ являются нормальными. Оптимизация моделей основана на процедуре настройки моделей в работе [14].

Во временных рядах Motion Capture Database сэмплирование во времени является равномерным, поэтому для порождения ситуации с переменным интервалом между наблюдениями было проведено маскирование данных по Бернулли: каждое наблюдение с вероятностью $\hat{p}$ независимо от других сохранялось в результирующей последовательности на своей позиции, иначе считалось, что в соответствующий момент времени наблюдений нет. Здесь и далее скрытые в результате маскирования наблюдения названы маскированными. Однако первые три наблюдения во всех временных рядах оставались нетронутыми. Каждая выборка временных рядов была разбита на обучающую и тестовую в отношении 9 к 1 по временным рядам (*out–of–sample*).

В рассматриваемых выборках были настроены модели вида (22) с помощью процедуры вариационного вывода на основе дивергенций α–Реньи. В проведённых экспериментах варьировалась частота пропусков и вид оптимизируемой оценки. Была сравнена стохастическая динамика с частично детерминированной, то есть с нулевой диффузией. Качество решения оценивалось по среднему логарифму правдоподобия и по квадратному корню среднеквадратической ошибки предсказания модели на тестовых выборках ($rmse$) в двух основных режимах: *экстраполяция* и *интерполяция*. Их отличия указаны в таблице 4. В режиме интерполяции оценивается способность модели восстанавливать пропущенные значения. Для аппроксимирующего процесса начальное значение скрытого состояния кодируется проходом рекуррентной нейронной сети по всему ряду с конца по непропущенным наблюдениям. В ре-

жиме экстраполяции оценивается способность модели восстановить вторую половину ряда по известной первой. В обеих частях ряда используются и оцениваются только непропущенные наблюдения. При экстраполяции рекуррентная нейронная сеть кодирует только известную половину непропущенных наблюдений, начиная с конца известной части.

Режим	**Значение** S	**Подпоследовательность для оценки** $rmse$	**Подпоследовательность для оценки логарифма правдоподобия**
интерполяция	$N' = S \leq N$ немаскированные	$(\mathbf{x}_{t_{i_l}})_{l=1}^{N''},\ N'' = N - N'$ маскированные	$(\mathbf{x}_{t_{i_l}})_{l=1}^{N'},\ N' \leq N$ немаскированные
экстраполяция	$\lfloor \frac{N'}{2} \rfloor = S \leq \lfloor \frac{N}{2} \rfloor$ немаскированные	$(\mathbf{x}_{t_{i_l}})_{l=1}^{N'''},\ N' - S = N''' \leq N - S,$ $t_{i_l} \in [t_{i_{S+1}}, t_{i_{N'}}]$ немаскированные	$(\mathbf{x}_{t_{i_l}})_{l=1}^{N'},\ N' \leq N$ немаскированные

Таблица 4. Режимы сравнения моделей.

Процедура обучения моделей выполнена в рамках алгоритма 4. Как указано в предыдущем разделе, модель процесса диффузии (22) оптимизирована с помощью стохастического (суб)градиентного метода на основе метода стохастической оптимизации Adam [48] с линейным отжигом оптимизируемого функционала в первую половину процесса настройки модели. На каждой итерации метода модель процесса диффузии (22) оптимизировалась на $N^{batch} = 5$ случайно выбранных рядах при $K = 10$ в течение $N^e = 400$ эпох. В течение всей процедуры оптимизации модели темп обучения η_e экспоненциально уменьшался, испытав на последней эпохе уменьшение почти в 10 раз по сравнению с первой эпохой настройки модели, по следующей формуле:

$$\eta_e = 0.01 \cdot 0.99425^{e-1},\ e = \overline{1, N^e} \text{ — номер эпохи.}$$

На практике при настройке модели процесса диффузии (22) был использован так называемый линейный отжиг, в котором на каждом этапе настройки модели производная Радона–Никодима между априорным распределением и аппроксимацией апостериорного распределения возводилась в степень β, и данная степень линейно увеличивалась с 0 до 1 включительно. Такая процедура, безусловно, лишает оптимизируемый функционал свойства быть нижней оценкой логарифма неполного правдоподобия при $\beta \neq 1$, но позволяет на начальных этапах лучше настроить модель под выборку [49]. В линейном отжиге оптимизируемый критерий

$\mathcal{L}_{\alpha}^{K}((\mathbf{x}_{t_i})_{i=1}^{N};\phi,\theta)$ из (34) заменён на $\mathcal{L}_{\alpha,\beta}^{K}((\mathbf{x}_{t_i})_{i=1}^{N};\phi,\theta)$:

$$\mathcal{L}_{\alpha,\beta}^{K}((\mathbf{x}_{t_i})_{i=1}^{N};\phi,\theta) = \frac{1}{1-\alpha}\mathbb{E}_{\hat{\mathrm{Q}}^{1:K}}\left[\ln\left(\frac{1}{K}\sum_{k=1}^{K}\left(\left(\prod_{i=1}^{N}\mathrm{p}(\mathbf{x}_{t_i}|\mathbf{z}^{k}(t_i);\theta)\left(\frac{\mathrm{p}(\mathbf{z}_0^k;\theta)}{\mathrm{q}(\mathbf{z}_0^k|(\mathbf{x}_{t_j})_{j=1}^{S};\phi)}\right)^{\beta}\right)^{1-\alpha}(M_{T,k})^{(\alpha-1)\beta}\right)\right)\right],$$

$$\beta = \min\left\{\frac{e-1}{199}, 1\right\},\ e = \overline{1, N^e}.$$

Монте–Карло оценка $\hat{\mathcal{L}}_{\alpha,\beta}^{K}((\mathbf{x}_{t_i})_{i=1}^{N};\phi,\theta)$ задаётся аналогично (37):

$$\hat{\mathcal{L}}_{\alpha,\beta}^{K}((\mathbf{x}_{t_i})_{i=1}^{N};\phi,\theta) = \frac{1}{1-\alpha}\ln\left(\frac{1}{K}\sum_{k=1}^{K}\left(\left(\prod_{i=1}^{N}\mathrm{p}(\mathbf{x}_{t_i}|\mathbf{z}^{k}(t_i);\theta)\left(\frac{\mathrm{p}(\mathbf{z}_0^k;\theta)}{\mathrm{q}(\mathbf{z}_0^k|(\mathbf{x}_{t_j})_{j=1}^{S};\phi)}\right)^{\beta}\right)^{1-\alpha}(M_{T,k})^{(\alpha-1)\beta}\right)\right),$$

$$\mathcal{L}_{\alpha,\beta}^{K}((\mathbf{x}_{t_i})_{i=1}^{N};\phi,\theta) = \mathbb{E}_{\hat{\mathrm{Q}}^{1:K}}\left[\hat{\mathcal{L}}_{\alpha,\beta}^{K}((\mathbf{x}_{t_i})_{i=1}^{N};\phi,\theta)\right].$$

Также аналогичная таблицам 1 и 2 таблица $\mathcal{L}_{\alpha,\beta}^{K}((\mathbf{x}_{t_i})_{i=1}^{N};\phi,\theta)$ для специальных значений α приведена ниже:

Значение α	**Оценка** $\mathcal{L}_{\alpha,\beta}^{K}((\mathbf{x}_{t_i})_{i=1}^{N};\phi,\theta)$
0	$\mathbb{E}_{\hat{\mathrm{Q}}^{1:K}}\left[\ln\left(\frac{1}{K}\sum_{k=1}^{K}\left(\prod_{i=1}^{N}\mathrm{p}(\mathbf{x}_{t_i}|\mathbf{z}^{k}(t_i);\theta)\left(\frac{\mathrm{p}(\mathbf{z}_0^k;\theta)}{\mathrm{q}(\mathbf{z}_0^k|(\mathbf{x}_{t_j})_{j=1}^{S};\phi)M_{T,k}}\right)^{\beta}\right)\right)\right]$
1	$\mathbb{E}_{\hat{\mathrm{Q}}^{1:K}}\left[\frac{1}{K}\sum_{k=1}^{K}\left(\sum_{i=1}^{N}\ln\left(\mathrm{p}(\mathbf{x}_{t_i}|\mathbf{z}^{k}(t_i);\theta)\right) - \beta\int_{0}^{T}\frac{1}{2}\left\|u(\mathbf{z}^{k}(t),t,\widetilde{\mathbf{x}})\right\|_2^2\mathrm{d}\,t\right)\right] - \beta\,\mathrm{KL}\left[\mathrm{q}(\mathbf{z}_0|(\mathbf{x}_{t_j})_{j=1}^{S};\phi)\middle\|\mathrm{p}(\mathbf{z}_0;\theta)\right]$
$+\infty$	$\mathbb{E}_{\hat{\mathrm{Q}}^{1:K}}\left[\min_{k\in\{1,\ \ldots,\ K\}}\left(\sum_{i=1}^{N}\ln\left(\mathrm{p}(\mathbf{x}_{t_i}|\mathbf{z}^{k}(t_i);\theta)\right) + \beta\left(\ln\left(\mathrm{p}(\mathbf{z}_0^k;\theta)\right) - \ln\left(\mathrm{q}(\mathbf{z}_0^k|(\mathbf{x}_{t_j})_{j=1}^{S};\phi)\right) - \ln\left(M_{T,k}\right)\right)\right)\right]$

Таблица 5. Виды вариационной оценки.

Логарифм правдоподобия оценен с помощью сэмплирования с выборкой по значимости при $K = 100$, значение $rmse$ для каждой конфигурации гиперпараметров эксперимента оценено с помощью $K = 100$ сэмплов последовательностей $(\mathbf{x}_{t_i}^k)_{i=1}^{N},\ k = \overline{1, K}$, сгенерированных из вариационного приближения. Результаты экспериментов приведены в таблицах 6, 7, 8, 9. Они демонстрируют заметный прирост в качестве при интерполяции маскированных (пропущенных) данных в случае стохастической динамики ($G(\cdot) \neq 0$) по сравнению с детерминированной динамикой ($G(\cdot) \equiv 0$) при, в среднем, 75% пропущенных наблюдений ($\hat{p} = 0.25$), однако при снижении доли пропусков детерминированная динамика с интерполяцией начинает справляться лучше. Аналогичная ситуация наблюдается в случае экстраполяции временного ряда. Также таблицы 6 и 8 демонстрируют практическую полезность применения дивергенций α–Реньи в процедуре вариационного вывода, так как большинство наилучших значений $rmse$ было получено при $\alpha \neq 1$.

7 Заключение

В работе были получены следующие основные результаты:

- разработана схема оптимизации нейронных сетей, параметризованных с помощью стохастических дифференциальных уравнений;
- разработана схема вариационного вывода с α–Реньи дивергенциями для вероятностных моделей, параметризованных с помощью стохастических дифференциальных уравнений;
- предложена модель описания временного ряда с неравномерным сэмплированием во времени;
- решена задача моделирования траектории движения человеческого скелета.

Результаты работы опубликованы в сборнике тезисов XXVII Международной научной конференции студентов, аспирантов и молодых учёных «Ломоносов–2020» [50].

Список литературы

1. Kingma Diederik P, Welling Max. Auto-encoding variational bayes // arXiv preprint arXiv:1312.6114. 2013.
2. Dinh Laurent, Krueger David, Bengio Yoshua. Nice: Non-linear independent components estimation // arXiv preprint arXiv:1410.8516. 2014.
3. Rezende Danilo Jimenez, Mohamed Shakir. Variational inference with normalizing flows // arXiv preprint arXiv:1505.05770. 2015.
4. Goodfellow Ian. NIPS 2016 tutorial: Generative adversarial networks // arXiv preprint arXiv:1701.00160. 2016.
5. Song Yang, Ermon Stefano. Generative modeling by estimating gradients of the data distribution // Advances in Neural Information Processing Systems. 2019. С. 11895–11907.
6. Variational approaches for auto-encoding generative adversarial networks / Mihaela Rosca, Balaji Lakshminarayanan, David Warde-Farley [и др.] // arXiv preprint arXiv:1706.04987. 2017.
7. Decomposed Adversarial Learned Inference / Alexander Hanbo Li, Yaqing Wang, Changyou Chen [и др.] // arXiv. 2020. С. arXiv–2004.
8. Khrulkov Valentin, Novikov Alexander, Oseledets Ivan. Expressive power of recurrent neural networks // arXiv preprint arXiv:1711.00811. 2017.

9. Wang Yixin, Blei David. Variational Bayes under Model Misspecification // Advances in Neural Information Processing Systems. 2019. С. 13357–13367.

10. Rubanova Yulia, Chen Tian Qi, Duvenaud David K. Latent Ordinary Differential Equations for Irregularly-Sampled Time Series // Advances in Neural Information Processing Systems. 2019. С. 5321–5331.

11. Brown ROBERT. Microscopical observations // Philos. Mag. 1828. Т. 4. С. 161–173.

12. Giles Mike, Glasserman Paul. Smoking adjoints: Fast monte carlo greeks // Risk. 2006. Т. 19, № 1. С. 88–92.

13. Yang Jichuan, Kushner Harold J. A Monte Carlo method for sensitivity analysis and parametric optimization of nonlinear stochastic systems // SIAM journal on control and optimization. 1991. Т. 29, № 5. С. 1216–1249.

14. Scalable Gradients for Stochastic Differential Equations / Xuechen Li, Ting-Kam Leonard Wong, Ricky TQ Chen [и др.] // arXiv preprint arXiv:2001.01328. 2020.

15. Deep residual learning for image recognition / Kaiming He, Xiangyu Zhang, Shaoqing Ren [и др.] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. С. 770–778.

16. Tran Quan, MacKinlay Andrew, Yepes Antonio Jimeno. Named entity recognition with stack residual lstm and trainable bias decoding // arXiv preprint arXiv:1706.07598. 2017.

17. Beyond finite layer neural networks: Bridging deep architectures and numerical differential equations / Yiping Lu, Aoxiao Zhong, Quanzheng Li [и др.] // arXiv preprint arXiv:1710.10121. 2017.

18. Haber Eldad, Ruthotto Lars. Stable architectures for deep neural networks // Inverse Problems. 2017. Т. 34, № 1. с. 014004.

19. Ruthotto Lars, Haber Eldad. Deep neural networks motivated by partial differential equations // arXiv preprint arXiv:1804.04272. 2018.

20. Neural ordinary differential equations / Tian Qi Chen, Yulia Rubanova, Jesse Bettencourt [и др.] // Advances in neural information processing systems. 2018. С. 6571–6583.

21. Ffjord: Free-form continuous dynamics for scalable reversible generative models / Will Grathwohl, Ricky TQ Chen, Jesse Bettencourt [и др.] // arXiv preprint arXiv:1810.01367. 2018.

22. Понтрягин Лев Семенович. Математическая теория оптимальных процессов. Гос. изд-во Физико-математической лит-ры, 1961.

23. Kushner Harold J, Yang Jichuan. A Monte Carlo method for sensitivity analysis and parametric optimization of nonlinear stochastic systems: the ergodic case // SIAM journal on control and optimization. 1992. Т. 30, № 2. С. 440–464.

24. Fleming Wendell H, McEneaney William M. Risk sensitive optimal control and differential games // Stochastic theory and adaptive control. Springer, 1992. С. 185–197.

25. Gobet Emmanuel, Munos Rémi. Sensitivity analysis using Itô–Malliavin calculus and martingales, and application to stochastic optimal control // SIAM Journal on control and optimization. 2005. Т. 43, № 5. С. 1676–1713.

26. Neural SDE: Stabilizing Neural ODE Networks with Stochastic Noise / Xuanqing Liu, Si Si, Qin Cao [и др.] // arXiv preprint arXiv:1906.02355. 2019.

27. Itô Kiyosi. 109. stochastic integral // Proceedings of the Imperial Academy. 1944. Т. 20, № 8. С. 519–524.

28. Protter Philip E. Stochastic differential equations // Stochastic integration and differential equations. Springer, 2005. С. 249–361.

29. Robbins Herbert, Monro Sutton. A stochastic approximation method // The annals of mathematical statistics. 1951. С. 400–407.

30. Øksendal Bernt. Stochastic differential equations // Stochastic differential equations. Springer, 2003. С. 65–84.

31. Higham Desmond J, Mao Xuerong, Stuart Andrew M. Strong convergence of Euler-type methods for nonlinear stochastic differential equations // SIAM Journal on Numerical Analysis. 2002. Т. 40, № 3. С. 1041–1063.

32. Blei David M, Kucukelbir Alp, McAuliffe Jon D. Variational inference: A review for statisticians // Journal of the American statistical Association. 2017. Т. 112, № 518. С. 859–877.

33. Titsias Michalis, Lázaro-Gredilla Miguel. Doubly stochastic variational Bayes for non-conjugate inference // International conference on machine learning. 2014. С. 1971–1979.

34. Li Yingzhen, Turner Richard E. Rényi divergence variational inference // Advances in Neural Information Processing Systems. 2016. С. 1073–1081.

35. Van Erven Tim, Harremos Peter. Rényi divergence and Kullback-Leibler divergence // IEEE Transactions on Information Theory. 2014. Т. 60, № 7. С. 3797–3820.

36. Burda Yuri, Grosse Roger, Salakhutdinov Ruslan. Importance weighted autoencoders // arXiv preprint arXiv:1509.00519. 2015.

37. Grünwald Peter D, Grunwald Abhijit. The minimum description length principle. MIT press, 2007.

38. Benedetto John J, Czaja Wojciech. Integration and modern analysis. Springer Science & Business Media, 2010.

39. Kunita Hiroshi. Stochastic Flows and Jump-Diffusions. Springer, 2019.

40. ANODEV2: A Coupled Neural ODE Framework / Tianjun Zhang, Zhewei Yao, Amir Gholami [и др.] // Advances in Neural Information Processing Systems. 2019. С. 5152–5162.

41. Bahlali Khaled, Djehiche Boualem, Mezerdi Brahim. On the stochastic maximum principle in optimal control of degenerate diffusions with Lipschitz coefficients // Applied mathematics and optimization. 2007. Т. 56, № 3. С. 364–378.

42. Karatzas I, Shreve S. Brownian Motion and Stochastic Calculus, second edtition Springer // New York. 1991.

43. Boué Michelle, Dupuis Paul [и др.]. A variational representation for certain functionals of Brownian motion // The Annals of Probability. 1998. Т. 26, № 4. С. 1641–1659.

44. Dupuis Paul, Ellis Richard S. A weak convergence approach to the theory of large deviations. John Wiley & Sons, 2011. Т. 902.

45. Durrett Richard, Durrett Richard. Brownian motion and martingales in analysis. Wadsworth Advanced Books & Software California, 1984.

46. Learning phrase representations using RNN encoder-decoder for statistical machine translation / Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre [и др.] // arXiv preprint arXiv:1406.1078. 2014.

47. CMU Graphics Lab Motion Capture Database. `http://mocap.cs.cmu.edu/`. Дата актуализации: 2020-05-23.

48. Kingma Diederik P, Ba Jimmy. Adam: A method for stochastic optimization // arXiv preprint arXiv:1412.6980. 2014.

49. beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework. / Irina Higgins, Loic Matthey, Arka Pal [и др.] // Iclr. 2017. Т. 2, № 5. с. 6.

50. Юдин Н. Адаптивный вариационный вывод // Сборник тезисов XXVII международной научной конференции студентов, аспирантов и молодых учёных «Ломоносов–2020». МАКС Пресс, 2020.

A Приложение

A.1 Результаты экспериментов

$\alpha :=$	0			0.5			1			$+\infty$		
$\hat{p} :=$	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75
субъект						$G(\cdot) \neq \mathbf{0}$						
83	0.39	0.36	0.55	0.43	0.42	0.44	**0.31**	0.34	0.37	0.53	0.51	0.38
79	0.36	0.25	**0.2**	0.39	0.35	0.35	**0.23**	0.31	0.2	0.29	0.37	0.26
138	**0.37**	0.37	0.44	0.49	0.36	**0.39**	0.53	0.48	0.41	0.36	0.41	0.55
69	0.45	0.46	**0.3**	0.32	0.33	0.32	0.39	0.42	0.37	0.39	0.35	0.42
80	0.44	0.42	0.36	0.48	0.57	0.37	0.39	0.34	0.27	0.29	0.4	**0.21**
субъект						$G(\cdot) \equiv \mathbf{0}$						
83	0.45	**0.29**	0.35	0.36	0.57	**0.32**	0.52	0.44	0.45	0.45	0.35	0.35
79	0.25	0.36	0.31	0.31	0.31	0.42	0.24	**0.24**	0.22	0.36	0.3	0.33
138	0.45	0.38	0.62	0.42	0.47	0.56	0.42	**0.36**	0.45	0.38	0.61	0.55
69	0.35	0.4	0.3	**0.28**	**0.29**	0.32	0.39	0.35	0.48	0.37	0.31	0.38
80	0.37	**0.28**	0.4	**0.21**	0.43	0.35	0.41	0.33	0.51	0.3	0.35	0.57

Таблица 6. *rmse* в режиме интерполяции, $K = 100$.

$\alpha :=$	0			0.5			1			$+\infty$		
$\hat{p} :=$	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75
субъект						$G(\cdot) \neq \mathbf{0}$						
83	0.13	0.28	0.29	0.14	0.18	0.26	0.13	0.21	0.27	0.07	0.11	0.21
79	0.25	0.13	0.89	0.27	0.66	0.72	0.30	0.61	0.94	0.24	0.64	0.91
138	0.28	0.41	0.89	0.29	0.52	0.49	0.33	0.36	0.73	0.34	0.36	0.69
69	0.36	0.9	1.42	0.42	0.8	1.34	0.49	0.95	1.24	0.37	0.85	1.16
80	0.99	1.85	3.91	1.33	1.9	3.1	1.15	2.45	3.58	1.41	2	3.86
субъект						$G(\cdot) \equiv \mathbf{0}$						
83	0.12	0.24	0.35	0.08	0.23	0.38	0.13	0.27	0.31	0.14	0.2	0.26
79	0.35	0.61	0.73	0.33	0.53	0.79	0.24	0.19	0.9	0.27	0.3	0.88
138	0.21	0.59	0.56	0.38	0.25	0.44	0.23	0.57	0.68	0.27	0.34	0.57
69	0.46	0.83	1.38	0.48	1.08	1.45	0.38	1.05	1.09	0.47	1	1.2
80	1.19	2.67	3.37	1.08	2.54	3.6	0.83	1.93	2.52	1.02	2.06	2.61

Таблица 7. $\times 10^4$ средний логарифм правдоподобия в режиме интерполяции, $K = 100$.

$\alpha :=$	0			0.5			1			$+\infty$		
$\hat{p} :=$	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75
субъект	$G(\cdot) \neq \mathbf{0}$											
83	0.38	0.4	0.42	0.34	0.42	0.34	0.38	0.44	0.35	**0.28**	0.58	0.41
79	0.45	0.21	0.27	0.35	**0.19**	0.35	0.23	0.27	0.31	0.35	0.19	0.38
138	0.41	0.47	0.44	0.6	0.62	**0.35**	0.55	0.61	0.53	0.52	0.46	0.5
69	0.34	0.42	0.37	0.52	0.4	0.34	0.49	0.31	0.34	0.42	0.37	0.3
80	0.3	0.42	0.42	0.33	0.31	**0.24**	**0.25**	0.3	0.42	0.29	0.37	0.42
субъект	$G(\cdot) \equiv \mathbf{0}$											
83	0.48	**0.31**	0.38	0.39	0.37	0.3	0.37	0.36	**0.29**	0.4	0.42	0.43
79	0.4	0.31	**0.22**	0.31	0.21	0.4	0.23	0.24	0.31	**0.21**	0.25	0.46
138	**0.4**	0.48	0.54	0.49	0.47	0.55	0.52	0.41	0.54	0.66	**0.38**	0.45
69	0.35	**0.31**	**0.25**	**0.3**	0.36	0.4	0.36	0.47	0.4	0.4	0.37	0.37
80	0.3	0.59	0.46	0.48	**0.26**	0.46	0.44	0.33	0.42	0.51	0.33	0.34

Таблица 8. $rmse$ в режиме экстраполяции, $K = 100$.

$\alpha :=$	0			0.5			1			$+\infty$		
$\hat{p} :=$	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75
субъект	$G(\cdot) \neq \mathbf{0}$											
83	0.09	0.24	0.28	0.15	0.14	0.43	0.07	0.21	0.37	0.16	0.13	0.31
79	0.26	0.41	0.73	0.09	0.71	0.92	0.25	0.53	0.8	0.33	0.56	0.93
138	0.07	0.55	0.65	0.14	0.44	1.07	0.13	0.32	0.07	0.25	0.65	0.73
69	0.53	0.7	1.09	0.43	0.73	1.37	0.41	0.92	1.23	0.36	0.79	1.46
80	1.03	1.9	3.55	1.18	2.31	3.65	1.02	2.22	3.24	1.08	2.03	3.43
субъект	$G(\cdot) \equiv \mathbf{0}$											
83	0.09	0.25	0.24	0.11	0.24	0.36	0.1	0.27	0.35	0.12	0.24	0.38
79	0.25	0.53	0.95	0.36	0.59	0.53	0.3	0.5	0.56	0.27	0.58	0.82
138	0.42	0.55	0.48	0.23	0.48	0.62	0.17	0.44	0.52	0.15	0.68	0.43
69	0.44	0.8	1.34	0.51	0.71	1.21	0.42	0.85	1.22	0.44	0.88	1.26
80	1.03	2.27	2.89	1.02	2.19	3.05	0.95	2.39	3.86	1.29	2.1	2.68

Таблица 9. $\times 10^4$ средний логарифм правдоподобия в режиме экстраполяции, $K = 100$.

A.2 Архитектура модели

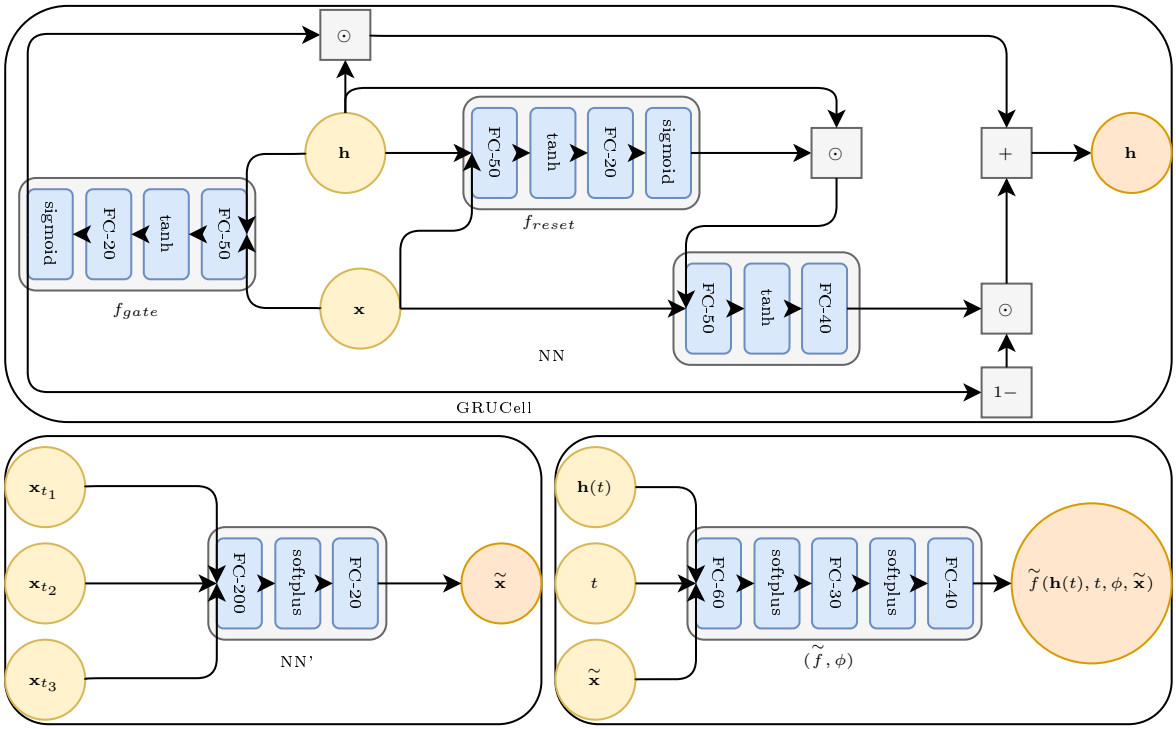

Рис. 2. Архитектура отображений в процедуре SequenceEncode.

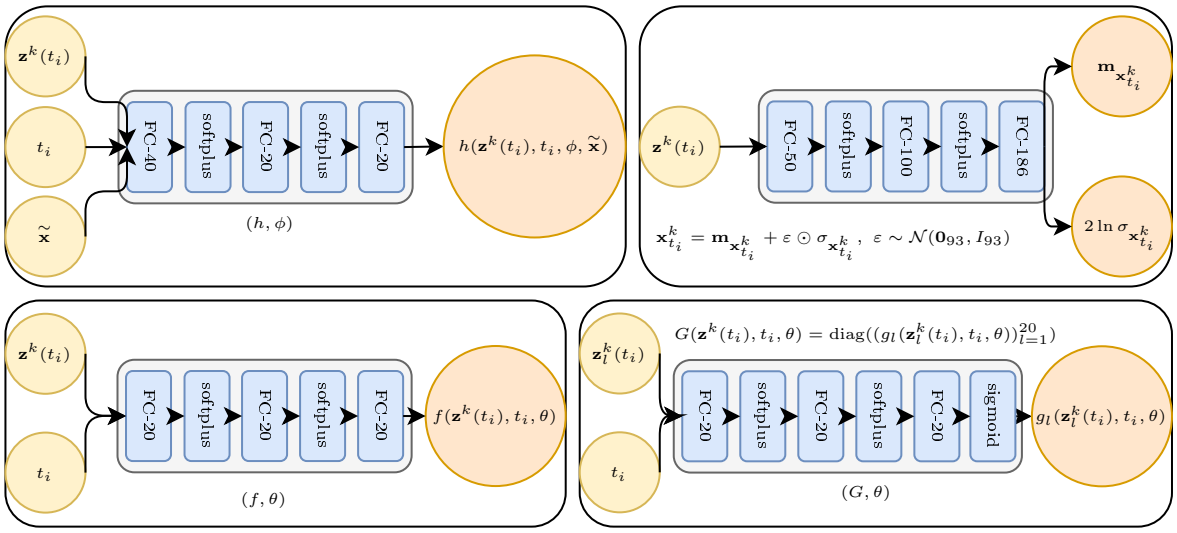

Рис. 3. Архитектура отображений в процедуре генерации последовательности.