

1. Основные понятия теории динамических систем.

Опр. 1 (Динамическая система). **динамической системой** называется пара из фазового пространства P (метрическое пространство или многообразие) и однопараметрической непрерывной или дискретной группы преобразований $P \times \mathbb{R} \rightarrow P$ или $P \times \mathbb{Z} \rightarrow P$, обозначаемой как $\phi^t(x), x \in P$. Для данной группы отображений должны выполняться следующие свойства:

- $\phi^0(x) = x$;
- $\phi^{t_1} \circ \phi^{t_2} = \phi^{t_1+t_2}$;
- ϕ^t дифференцируема по времени, определена и обратима для любых корректных значений t .

Опр. 2 (Траектория). **Траекторией**, проходящей через точку $x \in P$ называется множество $\{\phi^t(x) | t \in T\}$

В случае непрерывного времени динамическую систему можно задать уравнением следующего вида:

$$\dot{x} = F(x), \quad F(x) = \left. \frac{d\phi^t(x)}{dt} \right|_{t=0}.$$

2. Ляпуновский показатель. Вычисление ляпуновского показателя в случае анализа систем, заданных аналитически.

Погрешность предсказания любого временного ряда меняется по следующему закону:

$$\varepsilon_t = \varepsilon_0 e^{\lambda t},$$

где ε_0 – погрешность последнего известного значения ряда, ε_t – погрешность предсказания t шагов спустя, $\lambda = \text{const}$ – **старший показатель Ляпунова**. В случае регулярных рядов $\lambda < 0$, в случае хаотических – $\lambda > 0$.

Отсюда следует, что для хаотических рядов существует **горизонт прогнозируемости** – количество шагов T , такое, что после него погрешность любого предсказания превышает допустимый порог ε_{\max} : $T \sim \frac{1}{\lambda} \ln \frac{\varepsilon_{\max}}{\varepsilon_0}$.

Пусть рассматриваемый временной ряд соответствует динамической системе $\dot{x} = f(x)$. Зафиксируем точку $x(t)$, принадлежащую какой-либо траектории. Рассмотрим малое возмущение этой траектории $x(t) + \varepsilon u(t)$, также являющееся решением уравнения $\dot{x} = f(x)$. Подставляя разность двух рассматриваемых решений в уравнение, получаем следующее соотношение (считая, что f дифференцируема):

$$\begin{aligned}\varepsilon \dot{u} &= (x(t) + \varepsilon u(t)) - f(x(t)) = \frac{\partial f}{\partial x_i} \Big|_{x=x(t)} \cdot \varepsilon u(t) + O(\varepsilon), \\ \dot{u}(t) &= \nabla f|_{x(t)} \cdot u(t) = A(x(t)) \cdot u(t).\end{aligned}\tag{1}$$

Уравнение (1) называется линеаризацией динамической системы на траектории $x(t)$. Очевидно, что тождественный ноль является решением линеаризованной системы.

Если явный вид системы известен, что бывает чуть менее, чем никогда, то для нахождения спектра показателей Ляпунова можно применить алгоритм Бенеттина:

1. Рассмотрим совместно уравнение исходной динамической системы и n линеаризованных систем с ортонормированными начальными условиями. Удобно выбрать в качестве начальных условий для i -й системы вектор $e_i = (\delta_{ji})_{j=1}^n$. Полученная система в сумме состоит из $n \cdot (n + 1)$ скалярных уравнений.
2. Будем одновременно проводить численное интегрирование всех систем, используя метод Рунге-Кутты 4 порядка.
3. После каждого шага интегрирования будем производить ОГШ для решений линеаризованных систем, так, чтобы их решения оставались ортонормированными.
4. Также на каждом шаге вычисляем весь набор **объемных показателей Ляпунова**, пользуясь любым определением скалярного произведения:

$$\hat{\kappa}_j(t) = \frac{1}{t} \ln \sqrt{\det \begin{pmatrix} (u_1, u_1) & \dots & (u_1, u_j) \\ \vdots & \ddots & \vdots \\ (u_j, u_1) & \dots & (u_j, u_j) \end{pmatrix}}$$

5. В пределе при $t \rightarrow \infty$ находим значения объемных показателей Ляпунова.

Так как $\kappa_j = \sum_{i=1}^j \lambda_j$, то разности соседних κ_j дают показатели Ляпунова.

3. Ляпуновский показатель. Вычисление ляпуновского показателя по временному ряду. Метод аналога. Фрейм-разложение.

Рассмотрим временной ряд (возможно, многомерный) x_i . Тогда для данного ряда можно построить **ряд z -векторов**, каждый элемент которого определяется следующим образом:

$$\vec{z}_i = (x_i \ x_{i+1} \ \dots \ x_{i+d-1})^T,$$

$d = \text{const}$ – размерность вложения.

Построенный ряд можно использовать для оценки показателей Ляпунова. Для этого находятся пары близких z -векторов, после чего один из них рассматривается как возмущенная версия другого. Рассмотрим конкретные алгоритмы.

Метод аналога:

1. Для вектора «основной траектории» z_i находим его ближайшего соседа z_{j_1} ;
2. Для каждой такой пары (i, j_1) траектории z_{i+t} и z_{j_1+t} расходятся. В момент t_1 , когда норма разности между ними достигает некоторого порога $\varepsilon_2 = \text{const}$, вектор $z_{j_1+t_1}$ заменяется на другой вектор z_{j_2} , ближайший к z_{i+t_1} . Этот вектор также в какой-то момент t_2 заменяется на z_{j_3} и так далее.
3. Оценка старшего показателя Ляпунова вычисляется по следующей формуле:

$$\hat{\lambda}_1 = \frac{1}{t_n - t_0} \left(\ln \frac{\|z_{i+t_1} - z_{j_1+t_1}\|}{\|z_i - z_{j_1}\|} + \ln \frac{\|z_{i+t_2} - z_{j_2+t_2-t_1}\|}{\|z_{i+t_1} - z_{j_2}\|} + \dots + \ln \frac{\|z_{i+t_n} - z_{j_n+t_n-t_{n-1}}\|}{\|z_{i+t_{n-1}} - z_{j_n}\|} \right)$$

Метод фрейм-разложения.

4. Прогнозирование на основе кластеризации. Метод Уишарта.

Так как траектория системы прилегает к аттрактору и периодически проходит близко к самой себе в прошлом. Разумно предположить, что если начало

какого-то участка траектории близко к уже известному куску, называемому мотивом, то и продолжение этого участка совпадает с мотивом.

Как правило, в качестве мотивов выбираются центроиды кластеров в пространстве z -векторов, для чего необходимо уметь их кластеризовать. Однако рассматриваемая постановка задачи омрачается тем, что неизвестно ни количество кластеров, ни даже его порядок – десятки или миллионы, что не позволяет использовать многие из существующих алгоритмов кластеризации.

Одним из алгоритмов, которые можно использовать в данной задаче, является алгоритм Уишарта (Wishart), принимающий набор точек $V_n = \{x_i\}_{i=1}^n$ и два параметра k, h . Параметр k представляет собой количество рассматриваемых ближайших соседей и довольно слабо влияет на конечный результат. Параметр h является пороговым значением, участвующим в определении того, считается ли некоторый набор точек «значимым». Более формально, класс $C \subseteq V_n$ называется значимым по высоте h , если

$$\max_{x_i, x_j \in C} \left| \frac{k}{nW(d_k(x_i))} - \frac{k}{nW(d_k(x_j))} \right| \geq h,$$

где $W(r)$ – объем гипершара радиусом r , $d_k(x_i)$ – расстояние от точки x_i до ее k -го ближайшего соседа. При малых значениях h алгоритм возвращает множество кластеров, содержащих только одну точку, а при больших значениях h практически все точки не относятся ни к одному из найденных кластеров, то есть считаются межкластерным шумом.

Результатом работы алгоритма является набор меток кластеров w_i , где метка $w_j = 0$ обозначает межкластерный шум. Алгоритм состоит из следующих шагов:

1. Сортируем все точки по возрастанию $d_k(x_i)$. После этой операции в начале списка находятся так называемые «области сгущения» – точки, расположенные сравнительно близко к по крайней мере k другим точкам;
2. Рассматриваем граф G , вершины которого будут соответствовать объектам x_i . Изначально этот граф не содержит ни одной вершины; вершины добавляются по одной в соответствии с порядком, в котором отсортированы точки. Если в какой-то момент времени граф содержит n вершин, соответствующих n первым точкам, то точка x_{n+1} обрабатывается по следующим правилам:
 - (а) В граф добавляется новая вершина x_{n+1} (здесь и далее для простоты вершины графа обозначаются так же, как и исходные точки);

- (b) Добавленная вершина соединяется со всеми вершинами $x_j, j \leq n$, для которых выполняется условие $d(x_{n+1}, x_j) \leq d_k(x_j)$ (не наоборот, так, что нельзя утверждать, что количество соседей не превосходит k);
- (c) Если вершина x_{n+1} оказывается изолированной, то относим эту точку к новому кластеру;
- (d) Если вершина x_{n+1} не изолирована и все ее соседи принадлежат одному и тому же кластеру, то относим эту точку к этому же кластеру;
- (e) Если соседи x_{n+1} принадлежат к различным кластерам, среди которых нет нулевого кластера (обозначающего межкластерный шум) и ровно один кластер является значимым, то присваиваем к этому значимому кластеру как саму точку x_{n+1} , так и всех ее соседей;
- (f) В противном случае (то есть если соседи принадлежат к различным кластерам, причем или среди них есть нулевой, или количество значимых кластеров отлично от единицы), все кластера соседей вкупе с самой x_{n+1} присоединяются к межкластерному шуму.

При прогнозировании конец временного ряда сравнивается с началом каждого из обнаруженных мотивов. Если расстояние между ними не превышает некоторого порога, то оставшаяся часть мотива дает предсказание на соответствующее число шагов вперед. Так как количество полученных предсказаний может быть большим, то их необходимо агрегировать. Как правило, рассматривается кластеризация множества предсказаний. Если предсказания образуют один плотный кластер с небольшим количеством выбросов, то в качестве окончательного предсказания выбирается центроид этого кластера. Если же обнаруживается несколько разнесенных кластеров сравнимой мощности, то точка считается непрогнозируемой.

5. Плоскость энтропия-сложность.

Рассмотрим временной ряд x_t и соответствующий ему ряд из z -векторов размерности d :

$$\vec{z}_i = (x_i \ x_{i+1} \ \dots \ x_{i+d-1})^T.$$

Каждому такому вектору сопоставляется набор логических значений $f(z) \in$

$\Omega = \{0, 1\}^{d-1}$ по следующему правилу:

$$f : z = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_d \end{pmatrix} \mapsto \begin{pmatrix} [x_1 \leq x_2] \\ [x_2 \leq x_3] \\ \dots \\ [x_{d-1} \leq x_d] \end{pmatrix} \in \Omega.$$

Поскольку элементы временного ряда x_i можно рассматривать как наблюдения некоторого временного процесса, $f(z)$ является дискретной случайной величиной, обладающей каким-то распределением P .

Если рассматриваемый процесс является шумом, то P – это равномерное распределение U на множестве Ω . А хаотической динамической системе отвечает некое распределение $P \neq U$.¹

Случайное распределение можно описать в терминах энтропии и неравновесности. Дадим соответствующие определения:

Опр. 3 (Информация по Шеннону). Рассмотрим произвольное распределение P и событие A . Тогда **информацией Шеннона**, соответствующей этому событию, называется значение $I_A = f\left(\frac{1}{P(A)}\right)$, где f – некоторая функция, удовлетворяющая следующим свойствам:

- Функция f возрастает;
- Если события A, B независимы, то $I_{AB} = I_A + I_B$.

Можно показать, что этим условиям удовлетворяет только логарифмическая функция с произвольным основанием. В целях нормировки общепринято использование двоичного логарифма; единицей измерения информации в таком случае является бит. Таким образом, получаем:

$$I_A = \log_2 \frac{1}{P(A)} = -\log_2 P(A).$$

Опр. 4 (Энтропия по Шеннону). **Энтропией по Шеннону** называется среднее количество информации, получаемое в результате одного наблюдения:

$$H = \mathbb{E}_P I_\omega = \sum_i I_{\omega_i} p_i = - \sum_i p_i \log p_i.$$

¹Регулярной динамической системе также отвечает распределение, не являющееся равномерным, но к таким системам этот метод исследования обычно не применяется.

Известно, что энтропия минимальна и равна 0 в случае вырожденного распределения ($p_i = \delta_{ij}$) и максимальна в случае равномерного распределения. Также часто рассматривается **нормализованная энтропия**:

$$\bar{H}(P) = \frac{H(P)}{H_{\max}} = \frac{H(P)}{H(U)} \in [0, 1].$$

Опр. 5 (Сложность). Сложностью называется следующая величина:

$$S = Q[P, U] \cdot \bar{H}(P),$$

где неравновесность Q задается следующей формулой:

$$Q[P_1, P_2] = Q_0 \cdot \left(2H\left(\frac{P_1 + P_2}{2}\right) - H(P_1) - H(P_2) \right),$$

Q_0 – нормировочный коэффициент.

При правильном выборе нормировочных коэффициентов возможные положения точки (\bar{H}, S) на плоскости «энтропия-сложность» заключены между двумя параболой, $0 \leq H \leq 1, 0 \leq S \leq \frac{1}{2}$. При этом верхней части области соответствуют хаотические временные ряды, нижним областям – шум.

6. Инвариантная мера динамической системы.

Поскольку для хаотических систем сколь угодно малые отклонения в начальных условиях приводят к экспоненциально растущей ошибке предсказания, математический аппарат, описывающий эволюцию системы набором траекторий, оказывается не слишком полезным. Вместо этого используются понятия, взятые из теории вероятности.

Разобьем фазовое пространство X на ячейки, диаметры которых не превосходят $\varepsilon = \text{const}$. Для каждой ячейки вычислим долю времени, которое траектория системы проводит в этой ячейке за время T . Оказывается, что предельный результат при $T \rightarrow \infty$ не зависит от начальных условий и описывается некоторой вероятностной мерой на X . Эта мера называется **инвариантной мерой**, поскольку ее вид не меняется с течением времени.

Опр. 6 (Мера). **Мерой** на множестве X называется функция $\mu(A)$, определенная для некоторых подмножеств $A \subseteq X$ таким образом, что:

1. $\forall A : \mu(A) \geq 0$;
2. $\forall A, B : A \cap B = \emptyset \implies \mu(A \cup B) = \mu(A) + \mu(B)$;
3. $\mu(\emptyset) = 0$.²

Множества, на которых определено значение меры, называются **измеримыми**.

Если мера всего пространства $\mu(X) = 1$, то мера называется **вероятностной**.

Наиболее известна мера Лебега, соответствующая длине, площади, объему или их многомерному аналогу.

Рассмотрим некоторое вероятностное распределение на фазовом пространстве, заданное плотностью $p(x)$. Найдем вид этого распределения $q(y)$ при смене координат, заданной отображением $y = f(x)$. Отметим, что функция $q(y)$ также описывает вероятностное распределение на этом же фазовом пространстве после перемещения каждой точки x в $f(x)$. Из соображений сохранения вероятности получаем:

$$\begin{aligned} q(y) &= \sum_{x_i: f(x_i)=y} \frac{p(x_i)}{\left| \det \left(\frac{\partial f}{\partial x} \right) \right|} = \int_X \delta(f(x) - y) p(x_i) \frac{dy}{\left| \det \left(\frac{\partial f}{\partial x} \right) \right|} = \\ &= \int_X \delta(f(x) - y) p(x_i) dx. \end{aligned} \quad (2)$$

Теперь рассмотрим хаотическую динамическую систему с дискретным временем, заданную уравнением $x_{n+1} = f(x_n)$. Если в уравнение (2) в качестве начального распределения $p(x)$ подставить инвариантную меру, то после действия отображения вид распределения не должен измениться. Таким образом, получаем **уравнение Перрона-Фробениуса** для дискретного времени:

$$p(y) = \int \delta(f(x) - y) p(x) dx. \quad (3)$$

Аналогичное уравнение для системы $\dot{x} = F(x)$ с непрерывным временем может быть получено из (3). Рассмотрим отображение $\phi^\tau(x)$, ставящее в соответствие точке x точку $\hat{x}(\tau)$, где функция $\hat{x}(t)$ – решение задачи Коши $\dot{x} = F(x)$ с начальными условиями $x(0) = x$. Очевидно, что это преобразование также не

²Впрочем, это равенство следует из предыдущего.

должно менять вид плотности инвариантной меры. Следовательно, подставляя в (3), получаем:

$$p(y) = \int \delta(\phi^\tau(x) - y) p(x) dx.$$

Продифференцируем полученное уравнение по τ :³

$$0 = \int p(x) \nabla \delta(\phi^0(x) - y) \cdot d_\tau \phi^\tau(x) \big|_{\tau=0} dx.$$

Так как $\phi^\tau(x) = \hat{x}(\tau)$, то $d_\tau \phi^\tau(x) \big|_{\tau=0} = d_\tau \hat{x}(\tau) \big|_{\tau=0} = \dot{x}(0) = F(x)$, $\phi^0(x) = x$. В таком случае, по свойствам дельта-функции:

$$0 = \dots = \int p(x) \nabla \delta(x - y) \cdot F(x) dx = -\nabla(p(y)F(y)).$$

Таким образом, уравнение Перрона-Фробениуса свелось к хорошо изученному уравнению движения сжимаемой жидкости.

Теор. 1 (Крылова-Боголюбова). *Если существует хотя бы одно компактное множество A , инвариантное относительно $\phi^\tau(x)$, то для системы $\dot{x} = F(x)$ существует хотя бы одна вероятностная инвариантная мера μ .*

Теор. 2 (Эргодическая теорема (не мультипликативная)). *Пусть μ – инвариантная мера динамической системы, $g : \mathbb{R}^n \rightarrow \mathbb{R}$ – непрерывно измеримая функция на фазовом пространстве. Тогда для почти всех x по мере μ предельное среднее значение $g(x)$ равно теоретическому матожиданию:*

$$\forall t : \lim_{T \rightarrow +\infty} \frac{1}{T} \int_t^{t+T} g(x(\tau)) d\tau = \int g(x) \mu(dx) = \text{const}$$

Однако необходимо учесть, что понятие «почти всюду» для меры μ может значительно отличаться от такового для меры Лебега. Например, для системы $\dot{x} = -x$ («черная дыра») носитель меры состоит из одной точки $x = 0$; для всех остальных точек равенство может не соблюдаться.

Теор. 3 (Теорема Пуанкаре о возвращении множеств). *Пусть A – измеримое множество, инвариантная мера которого больше нуля. Тогда $\exists t > 1 : \mu(A \cap \phi^t(A)) > 0$. Иными словами, траектории, начинающиеся в данном множестве, будут бесконечно много раз возвращаться в это же множество.*

³В общем случае $p(y)$ может зависеть от времени, в таком случае в левой части получим $\partial_\tau p(y, t + \tau)$

Теор. 4 (Теорема Пуанкаре о возвращении траекторий). *Почти все точки x по мере μ устойчивы по Пуассону. Под устойчивостью по Пуассону понимается следующее свойство: для любой окрестности $U(x) \forall T : \exists t > T : \phi^t(x) \in U$, т.е. любая траектория бесконечно много раз возвращается в окрестность своей начальной точки.*

7. Энтропия Колмогорова-Синяя.

Опр. 7 (Энтропия Колмогорова-Синяя). Рассмотрим динамическую систему с дискретным временем $x_{n+1} = f(x_n)$ или с дискретизованным непрерывным временем $x(t + \tau) = \phi^\tau(x(t))$. Разобьем фазовое пространство на непересекающиеся множества A_i , такие, что $\text{diam } A_i < \varepsilon$ по любой метрике, например, Евклидовой.

Тогда введем следующую последовательность разбиений:

$$A_{i_1 i_2 \dots i_k} = \bigcap_{j=1}^k f^{1-j} (A_{i_j}),$$

то есть отнесем точку x к множеству $A_{i_1 i_2 \dots i_k}$, если изначально она находится в множестве A_{i_1} , после одного временного шага в множестве A_{i_2} и так далее до A_{i_k} . Заметим, что некоторые из построенных множеств (или даже подавляющее большинство) могут быть пустыми.

Для каждого разбиения глубины k подсчитаем энтропию Шеннона, соответствующую инвариантной мере системы μ :

$$H(k) = - \sum_{i \in \{1, \dots, N\}^k} \mu(A_{*i}) \log \mu(A_{*i})$$

Тогда **метрической энтропией** или **энтропией Колмогорова-Синяя** называется предельное приращение $H(k)$ при росте k :

$$K = \lim_{\varepsilon \rightarrow 0} \lim_{k \rightarrow \infty} (H(k+1) - H(k)) = \lim_{\varepsilon \rightarrow 0} \lim_{k \rightarrow \infty} \frac{H(k)}{k}$$

Для хаотических систем $K > 0$, что означает, что любое конечное количество информации о системе в начальный момент времени постепенно перестает быть хоть сколько-нибудь полезным в силу нарастания неопределенности. О таких системах говорят, что они производят информацию. Для регулярных динамических систем $K = 0$.