

ДИНАМИКА РАЗОБУЧЕНИЯ В НЕЙРОННОЙ СЕТИ ХОПФИЛДА

А.Ю.Плахов

В статье изучается динамика сходящегося итерационного алгоритма разобучения, предложенного в работах [7, 8]. В термодинамическом пределе получена самосогласованная система уравнений спектральной динамики синаптической матрицы. Найдена (изменяющаяся в процессе итерирования) сила разобучения, оптимизирующая сходимость алгоритма к проектор-матрице. Для оптимального разобучения определена динамика спектра синаптической матрицы.

1. ВВЕДЕНИЕ

За последнее десятилетие появилось большое количество работ по моделям ассоциативной памяти в нейронных сетях Хопфилда [1–10]. Значительные усилия были направлены на разработку обучающих правил, т.е. правил, по которым вычисляются коэффициенты синаптических связей. Многие важные результаты были получены для простого локального правила Хебба [1], запоминающего p образов (прототипов) за p последовательных шагов-предъявлений. Однако это правило позволяет запомнить лишь сравнительно небольшое количество некоррелированных образов ($p < 0.14N$, где N — число нейронов в сети [2] и, кроме того, восстановленные в процессе сетевой динамики конфигурации могут значительно отличаться от запомненных прототипов.

В последнее время были разработаны эффективные итерационные обучающие алгоритмы, состоящие в многократном предъявлении информации для запоминания [3–5]. Алгоритм, предложенный Дидерихом и Оппером, запоминает без ошибок $p < N$ произвольных образов [3, 6]. В работе [5] показано, что для точного запоминания каждый образ достаточно предъявить не более $\log N$ раз.

В работе [7] был предложен сходящийся алгоритм разобучения, функционирующий без многократного предъявления образов: после формирования синаптической матрицы по правилу Хебба сеть дообучается путем итерационной стохастической коррекции синаптических сетей. Согласно [7], корректирующий алгоритм состоит в следующем. На каждом шаге алгоритма случайно и независимо выбирается состояние сети $S = (S_1, \dots, S_N)$ (S_i принимают значений 1 и -1 с вероятностью 1/2),

затем вычисляются локальные поля $h_i = \sum_{j=1}^N J_{ij} S_j$, и коэффициенты синаптической связи модифицируются согласно

$$J_{ij} \rightarrow J_{ij} - \frac{\epsilon}{N} h_i h_j, \quad i, j = 1, \dots, N \quad (\epsilon > 0). \quad (1)$$

Таким образом, на каждом шаге итерации происходит разобучение ("обучение наоборот") конфигурации локальных полей, отвечающих случайному состоянию S . Алгоритм стартует с матрицы хеббовских связей

$$J_{ij}^H = \frac{1}{N} \sum_{\mu=1}^p \xi_i^\mu \xi_j^\mu, \quad i, j = 1, \dots, N, \quad (2)$$

где $\xi^\mu = (\xi_1^\mu, \dots, \xi_N^\mu)$, $\mu = 1, \dots, p$ ($\xi_i^\mu = \pm 1$) — образы для запоминания. В общем случае сила разобучения ϵ может изменяться в процессе итерирования. Если максимальное значение ϵ не превосходит некоторой критической величины $\epsilon_c > 0$, зависящей только от свойств набора образов, то (нормированная) синаптическая матрица асимптотически сходится к проектор-матрице P_{ij} , которая определяется как матрица ортогонального проектирования в N -мерном евклидовом пространстве на подпространство, натянутое на вектора ξ^1, \dots, ξ^p . Хорошо известно [9, 10], что матрица P_{ij} обеспечивает точное хранение в памяти набора $p < N$ образов; тем самым в результате действия алгоритма в асимптотике реализуется точное запоминание набора образов.

Сходимость синаптической матрицы к проектор-матрице доказана в [7, 8] для постоянного ϵ . Доказательство сходимости может быть легко распространено на случай ϵ , изменяющегося в процессе итерирования.

В настоящей работе изучается динамика процесса разобучения. Во втором разделе в термодинамическом пределе получена самосогласованная система уравнений спектральной динамики для синаптической матрицы и приводится эвристическое правило нахождения границ спектра. В следующем разделе вычисляется (зависящая от времени) сила разобучения, оптимизирующая сходимость алгоритма. Для оптимального разобучения определены скорость сходимости к проектор-матрице, а также динамика спектра матрицы синаптических связей. Статью завершает обсуждение перспектив дальнейшего исследования.

2. ВЫВОД УРАВНЕНИЯ СПЕКТРАЛЬНОЙ ДИНАМИКИ

Обозначим $J^{(n)}$ матрицу связности после n шагов разобучения и ϵ_n — силу разобучения на n -м шаге. Тогда итерационная формула (1) принимает вид

$$J_{ij}^{(n)} = J_{ij}^{(n-1)} - \frac{\epsilon_n}{N} h_i^{(n)} h_j^{(n)},$$

где

$$h_i^{(n)} = \sum_{j=1}^N J_{ij}^{(n-1)} S_j^{(n)},$$

или в матричной форме

$$\mathbf{J}^{(n)} = \mathbf{J}^{(n-1)} - \varepsilon_n \mathbf{J}^{(n-1)} \mathbf{P}^{(n)} \mathbf{J}^{(n-1)} \quad (3)$$

с начальным условием $\mathbf{J}^{(0)} = \mathbf{J}^H$. Здесь $\mathbf{P}^{(n)}$ обозначает матрицу с элементами $N^{-1} S_i^{(n)} S_j^{(n)}$, где $\vec{S}^{(n)}$ — вектор состояния сети, случайно выбранной на n -шаге.

Удобно определить следующую формулу Грина

$$u^{(n)}(z) = \left\langle \frac{1}{N} \text{Tr} R_z^{(n)} \right\rangle_S,$$

где

$$R_z^{(n)} = \mathbf{J}^{(n)} (\mathbf{I} - z \mathbf{J}^{(n)})^{-1},$$

\mathbf{I} — единичная матрица, $\langle \rangle_S$ обозначает усреднение по всевозможным случайным выборкам $\{\vec{S}^{(n)}, n = 1, 2, \dots\}$. Эта функция определена для всех комплексных значений z , за исключением λ_i^{-1} , где λ_i обозначает ненулевые собственные значения матрицы $\mathbf{J}^{(n)}$.

Используя $u^{(n)}(z)$, можно вычислить среднюю плотность ненулевых собственных значений $\mathbf{J}^{(n)}$ по формуле

$$\rho^{(n)}(\lambda) = \pi^{-1} \lambda^{-2} \text{Im} u^{(n)}(\lambda^{-1} + i0), \quad \lambda \neq 0. \quad (4)$$

Действительно, $u^{(n)}(z)$ может быть представлена в виде

$$u^{(n)}(z) = \left\langle \frac{1}{N} \sum_i \frac{\lambda_i}{1 - z \lambda_i} \right\rangle_S = \int_{-\infty}^{\infty} \frac{\lambda}{1 - z \lambda} \rho^{(n)}(\lambda) d\lambda.$$

Подставляя в правую часть этой формулы $x^{-1} + i0$ вместо z и используя представление для дельта-функции $\delta(\lambda) = \pi^{-1} \text{Im} [(\lambda - i0)^{-1}]$, получаем

$$\pi^{-1} \text{Im} u^{(n)}(x^{-1} + i0) = \int \delta(\lambda^{-1} - x^{-1}) \rho^{(n)}(\lambda) d\lambda = x^2 \rho^{(n)}(x).$$

Выведем в термодинамическом пределе уравнение динамики для функции Грина. Для этого оценим приращение $\delta u^{(n)}(z) = u^{(n)}(z) - u^{(n-1)}(z)$ за один шаг итерации. Прежде всего заметим, что

$$R_z^{(n)} - R_z^{(n-1)} = -\varepsilon_n \left[1 + \frac{\varepsilon_n z}{N} \sum_{ij} (R_z^{(n-1)} J^{(n-1)})_{ij} S_i^{(n)} S_j^{(n)} \right]^{-1} R_z^{(n-1)} P^{(n)} R_z^{(n-1)}.$$

Это равенство можно проверить, умножая обе его части на $\mathbf{I} - z\mathbf{J}^{(n)}$ и используя тождество

$$\mathbf{P}^{(n)}\mathbf{M}\mathbf{P}^{(n)} = \left(\frac{1}{N} \sum_{ij} M_{ij} S_i^{(n)} S_j^{(n)} \right) \cdot \mathbf{P}^{(n)},$$

(матрица \mathbf{M} произвольна). Тогда выражение для $\delta u^{(n)}(z)$ приобретает вид

$$\delta u^{(n)}(z) = -\frac{1}{N} \left\langle \frac{A}{B} \right\rangle_S, \quad (5)$$

где

$$A = \text{Tr} (R_z^{(n-1)} P^{(n)} R_z^{(n-1)}), \quad B = \varepsilon_n^{-1} + \frac{z}{N} \sum_{ij} (R_z^{(n-1)} J^{(n-1)})_{ij} S_i^{(n)} S_j^{(n)}.$$

Далее, используя перестановочное свойство следа, для A получаем

$$A = \text{Tr} \left[(R_z^{(n-1)})^2 P^{(n)} \right] = \frac{1}{N} \sum_{ij=1}^N (R_z^{(n-1)})_{ij}^2 S_i^{(n)} S_j^{(n)}. \quad (6)$$

В силу случайного выбора векторов $\vec{S}^{(n)}$ величина A самоусредняется при $N \rightarrow \infty$. Так как $\langle S_i^{(n)} S_j^{(n)} \rangle_S = \delta_{ij}$, из (6) находим

$$\langle A \rangle_S = \frac{1}{N} \sum_{ij=1}^N \left\langle (R_z^{(n-1)})_{ij}^2 \right\rangle_S \delta_{ij} = \left\langle \frac{1}{N} \text{Tr} (R_z^{(n-1)})^2 \right\rangle_S,$$

или, с учетом тождества $(R_z^{(n-1)})^2 = (d/dz) R_z^{(n-1)}$,

$$\langle A \rangle_S = \frac{du^{(n-1)}(z)}{dz}. \quad (7)$$

В пределе $N \rightarrow \infty$, B также является самоусредняющейся величиной, причем усреднение дает

$$\langle B \rangle_S = \varepsilon_n^{-1} + \left\langle N^{-1} \text{Tr} [z R_z^{(n-1)} J^{(n-1)}] \right\rangle_S,$$

и с использованием тождества $z R_z^{(n-1)} J^{(n-1)} = R_z^{(n-1)} - R_0^{(n-1)}$ имеем

$$\langle B \rangle_S = \varepsilon_n^{-1} + u^{(n-1)}(z) - u^{(n-1)}(0). \quad (8)$$

В итоге, суммируя (5), (7), (8), приходим к следующему выражению для $\delta u^{(n)}(z)$

$$\delta u^{(n)}(z) = -\frac{1}{N} \left[\varepsilon_n^{-1} + u^{(n-1)}(z) - u^{(n-1)}(0) \right]^{-1} \frac{du^{(n-1)}(z)}{dz}.$$

Переходя к пределу $n, N \rightarrow \infty$ при конечном $t = n/N$ и производя замену $\varepsilon_n \rightarrow \varepsilon(t)$ и $u^{(n)}(z) \rightarrow u(z, t)$, получаем дифференциальное уравнение

$$\frac{\partial u(z, t)}{\partial t} = -\frac{1}{y(t) + u(z, t)} \frac{\partial u(z, t)}{\partial z}, \quad (9)$$

где

$$y(t) = \varepsilon^{-1}(t) - u(0, t). \quad (10)$$

Непрерывная переменная t здесь интерпретируется как текущее время разобучения.

Начальное условие для уравнения (9) определяется из соотношения

$$u(z, 0) = \int_{-\infty}^{\infty} \frac{\lambda}{1 - z\lambda} \rho^H(\lambda) d\lambda,$$

$\rho^H(\lambda)$ обозначает плотность собственных значений матрицы Хебба \mathbf{J}^H (2). Как известно (см., например, [11, 12]), в случае несмещенных образов (т.е. в случае, когда ξ_i^{μ} независимы и принимают значения ± 1 с вероятностью $1/2$) в пределе $p, N \rightarrow \infty, p/N = \alpha < 1$ плотность $\rho^H(\lambda)$ равна

$$\rho^H(\lambda) = (1 - \alpha)\delta(\lambda) + (2\pi\lambda)^{-1} \sqrt{4\alpha - (\lambda - 1 - \alpha)^2}.$$

Опуская технические подробности, приведем окончательное выражение для $u(z, 0)$

$$u(z, 0) = \frac{1}{2z^2} \left[1 - (1 + \alpha)z + (1 - \alpha)\sqrt{(z - x_1)(z - x_2)} \right], \quad (11)$$

$$x_{1,2} = (1 \pm \sqrt{\alpha})^{-2}.$$

В процессе итерирования (3), как можно показать, сохраняя нулевое собственное значение кратности $N - p$, соответствующее сингулярной части спектра $(1 - \alpha)\delta(\lambda)$. Оставшиеся p ненулевых собственных значений составляют в пределе $p, N \rightarrow \infty$ непрерывную часть спектра, плотность которой $\rho(\lambda, t)$ можно определить, переходя к пределу $n, N \rightarrow \infty, n/N = t$ в уравнении (4):

$$\rho(\lambda, t) = \pi^{-1} \lambda^{-2} \operatorname{Im} u(\lambda^{-1} + i0, t), \quad \lambda \neq 0. \quad (12)$$

Таким образом, система уравнений (9), (10) с начальным условием (11) вместе с соотношением (12) определяют эволюцию спектральной плотности синалтической матрицы $(1 - \alpha)\delta(\lambda) + \rho(\lambda, t)$.

В заключение этого раздела мы получим удобное для вычисления эвристическое правило определения непрерывной части спектра (см. также

[11]). Заметим, что функция $u(x^{-1}, t) = \int x \lambda (x - \lambda)^{-1} \rho(\lambda, t) d\lambda$ действительна и монотонно убывает на интервалах действительной оси, дополнительных к непрерывной части спектра (монотонное убывание проверяется непосредственным дифференцированием по x). Поэтому обратная к ней функция $x^{-1}(u, t)$ также монотонно убывает и область ее значений совпадает с дополнением к непрерывной части спектра.

Найдем функцию $x^{-1}(u, t)$. Для этого перепишем уравнение (9) в терминах $x(u, t)$:

$$\frac{\partial x(u, t)}{\partial t} = \frac{1}{y(t) + u}.$$

После интегрирования по t получаем

$$x(u, t) = F(u) + \int_0^t \frac{d\tau}{y(\tau) + u}, \quad (13)$$

где

$$F(u) = \frac{1}{u} \left(\sqrt{u + \left(\frac{1-\alpha}{2}\right)^2} - \frac{1+\alpha}{2} \right)$$

— функция, обратная к $u(x, 0)$ (11). $u(x, 0)$ определена на интервалах действительной оси $(-\infty, x_1) \cup (x_2, +\infty)$ и взаимно однозначно отображает их на $(u_1, 0) \cup (0, u_2)$, $u_{1,2} = \mp \sqrt{\alpha}(1 \mp \sqrt{\alpha})^2$, причем $\lim_{x \rightarrow \pm\infty} u(x, 0) = 0$. Поэтому функция $F(u)$ определена на интервале (u_1, u_2) (при $u = 0$ полагаем $F(0) = \infty$). Суммируя сказанное, сформулируем правило вычисления спектра. Определим интервалы монотонного убывания функции

$$x^{-1}(u, t) = \left(F(u) + \int_0^t \frac{d\tau}{y(\tau) + u} \right)^{-1}, \quad u \in (u_1, u_2),$$

и затем найдем множество ее значений на этих интервалах. Дополнение к этому множеству составляет непрерывную часть спектра.

3. ДИНАМИКА ОПТИМАЛЬНОГО РАЗОБУЧЕНИЯ

В этом разделе вычисляется (зависящая от времени) сила разобучения, оптимизирующая сходимость разобучающего алгоритма, а также определяется динамика спектра синаптической матрицы при оптимальном разобучении.

Прежде всего нам нужно выбрать меру близости между матрицами \mathbf{J} и \mathbf{P} . Воспользуемся тем, что нормировка матрицы синаптических связей оставляет сетевую динамику инвариантной, и рассмотрим в качестве

меры близости евклидово расстояние между нормированными матрицами $\|\mathbf{J}\|^{-1}\mathbf{J}$ и $\|\mathbf{P}\|^{-1}\mathbf{P}$, т.е.

$$D(J, P) = \left\| \|\mathbf{J}\|^{-1}\mathbf{J} - \|\mathbf{P}\|^{-1}\mathbf{P} \right\|, \quad (14)$$

где $\|\dots\|$ обозначает евклидову норму,

$$\|\mathbf{J}\| = \left(\sum_{ij=1}^N J_{ij}^2 \right)^{1/2} = (\text{Tr} \mathbf{J}^2)^{1/2}.$$

Определим задачу оптимизации следующим образом: выбрать значения силы разобучения $\varepsilon_1, \dots, \varepsilon_n$ таким образом, чтобы минимизировать расстояние $D(\mathbf{J}^{(n)}, \mathbf{P})$ после n шагов итерации.

Мы решаем эту задачу в пределе $n, N \rightarrow \infty, n/N = T > 0$. Используя (14), преобразуем величину $D = D(\mathbf{J}^{(n)}, \mathbf{P})$ к более удобному виду:

$$\begin{aligned} D^2 &= \text{Tr} \left(\|\mathbf{J}^{(n)}\|^{-2} (\mathbf{J}^{(n)})^2 + \|\mathbf{P}\|^{-2} \mathbf{P}^2 - 2 \|\mathbf{J}^{(n)}\|^{-1} \|\mathbf{P}\|^{-1} \mathbf{J}^{(n)} \mathbf{P} \right) = \\ &= 2 - \frac{\frac{2}{N} \text{Tr}(\mathbf{J}^{(n)} \mathbf{P})}{\left(\frac{1}{N} \text{Tr} \mathbf{P}^2 \right)^{1/2} \left(\frac{1}{N} \text{Tr} (\mathbf{J}^{(n)})^2 \right)^{1/2}}. \end{aligned}$$

Поскольку \mathbf{P} — проектор на p -мерное подпространство, для него справедливо соотношение $\text{Tr} \mathbf{P}^2 = \text{Tr} \mathbf{P} = p$. Кроме того, синаптическая матрица $\mathbf{J}^{(n)}$ удовлетворяет равенству $\mathbf{J}^{(n)} \mathbf{P} = \mathbf{J}^{(n)}$, которое может быть доказано индукцией по n с использованием формул (2), (3). С учетом вышесказанного, а также принимая во внимание очевидные равенства

$$\left\langle \frac{1}{N} \text{Tr} \mathbf{J}^{(n)} \right\rangle_s = u(0, T)$$

и

$$\left\langle \frac{1}{N} \text{Tr} (\mathbf{J}^{(n)})^2 \right\rangle_s = \frac{\partial u}{\partial z}(0, T),$$

получаем выражение для D :

$$D^2 = 2 - 2u(0, T) \left(\alpha \cdot \frac{\partial u}{\partial z}(0, T) \right)^{-1/2}.$$

Это выражение можно записать в терминах обратной функции $x(u, t)$ (13) в виде

$$D^2 = 2 - \frac{2}{\sqrt{\alpha}} \sqrt{u^2 \frac{\partial x}{\partial u}(u, T)}.$$

Здесь обозначено $u = u(0, T)$. В итоге, с учетом проделанных преобразований, задача оптимизации сводится к следующему:

1) Найти функцию $y(t)$, $t \in [0, T]$, максимизирующую величину

$$E = u^2 \left[F'(u) - \int_0^t \frac{dt}{(y(t) + u)^2} \right], \quad (15)$$

где u определяется из условия

$$x(u, T) = F(u) + \int_0^t \frac{dt}{y(t) + u} = 0. \quad (16)$$

2) Вычислить силу разобучения по формуле $\varepsilon(t) = (y(t) + v(t))^{-1}$, где функция $v(t) = u(0, t)$ определяется из условия

$$F(v(t)) + \int_0^t \frac{d\tau}{y(\tau) + v(t)} = 0.$$

Вначале покажем, что $y(t)$ должна быть тождественно постоянной. Предположив противное, подберем число y таким образом, чтобы

$$\frac{T}{y + u} = \int_0^T \frac{dt}{y(t) + u} \quad (17)$$

и рассмотрим разобучение "со звездочкой", соответствующее тождественно постоянной функции $y^*(t) \equiv y$. Для этого разобучения, в силу (17), имеем $x^*(u, T) = 0$, поэтому $u = u^*$. Кроме того, согласно неравенству Коши-Буняковского,

$$\int_0^T \frac{dt}{(y(t) + u)^2} > \frac{1}{T} \left(\int_0^T \frac{dt}{y(t) + u} \right)^2 = \frac{T}{(y + u)^2}.$$

В результате получаем $E^* > E$, что противоречит предположению о том, что функция $y(t)$ максимизирует E .

Постоянная $y(t) \equiv y$ в силу (16) может быть представлена как

$$y = -u - \frac{T}{F(u)}, \quad (18)$$

и с учетом (18) величина E принимает следующий вид

$$E = u^2 [F'(u) - T^{-1} F^2(u)]. \quad (19)$$

Анализ этой формулы показывает, что величина E (19) достигает максимума в точке u , служащей единственным решением уравнения

$$T = -[4u + (1 - \alpha)^2] \cdot F(u) \quad (20)$$

(соответствующие выкладки мы опускаем ввиду их громоздкости). С учетом (18) и (20) находим

$$y = 3u + (1 - \alpha)^2.$$

Отсюда с использованием (10) получаем

$$\varepsilon(t) = [(1 - \alpha)^2 + 3u + v]^{-1}, \quad (21)$$

где $v = u(0, t)$. Используя равенство $x(v, t) = F(v) + t/(y + v) = 0$, выразим t через v :

$$t = -[(1 - \alpha)^2 + 3u + v] \cdot F(v). \quad (22)$$

Когда t возрастает от 0 до T , v монотонно изменяется от α до u .

Формулы (20)–(22) дают решение задачи оптимизации разобучения на конечном промежутке времени T . Алгоритм построения оптимальной функции $\varepsilon(t)$, $t \in [0, T]$ следующий: для данного T по формуле (20) определяем значение u и затем подставляем его в формулы (21), (22), которые неявно задают $\varepsilon(t)$. В случае неограниченного времени разобучения $T \rightarrow \infty$ $u \rightarrow 0$, и (21), (22) принимают вид

$$t = -[(1 - \alpha)^2 + v] \cdot F(v),$$

$$\varepsilon(t) = [(1 - \alpha)^2 + v]^{-1}.$$

Когда свободный параметр v пробегает значения от α до 0, время t изменяется от 0 до ∞ и $\varepsilon(t)$ монотонно растет от $\varepsilon(0) = (1 - \alpha + \alpha^2)^{-1}$ до $\varepsilon(\infty) = (1 - \alpha)^{-2}$. График функции $\varepsilon(t)$ для трех значений параметра α приведен на рис.1. Интересно отметить, что $\varepsilon(0)$ примерно одинаково для всех значений $0 < \alpha < 1$ и варьирует между 1 и 1.33, в то время как предельное значение $\varepsilon(\infty)$ стремится к бесконечности, когда $\alpha \rightarrow 1$. При $t \gg 1$ асимптотическое поведение $\varepsilon(t)$ имеет вид

$$\varepsilon(t) \simeq (1 - \alpha)^{-2} - \alpha(1 - \alpha)^{-2}t^{-1}.$$

На рис.2 изображен график зависимости от времени расстояния D между текущей матрицей связности и проектор-матрицей. Отметим некоторое увеличение D в начальный период итерирования (т.е. в начале оптимального разобучения синаптическая матрица несколько "удаляется" от проектор-матрицы). Вычисления показывают, что этот эффект "ухудшения" синаптической матрицы проявляется при значениях $\alpha > \alpha_c$, где

$\alpha_c = 2 - \sqrt{3} \approx 0.27$. Асимптотика D при больших значениях t имеет вид $D \approx \sqrt{\alpha}/\sqrt{t}$, $t \rightarrow \infty$.

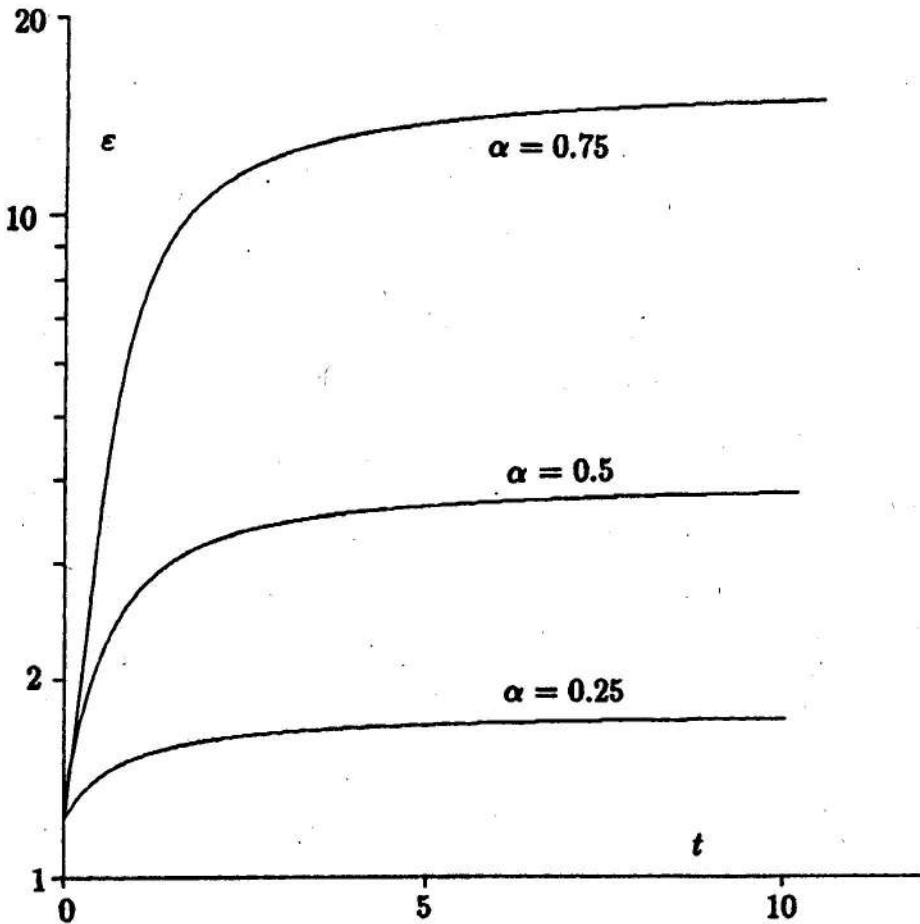


Рис. 1.

Рассмотрим динамику спектра синаптической матрицы при оптимальном разобучении с неограниченным временем. Мы будем иметь дело с нормированной матрицей $c\mathbf{J}$, причем коэффициент $c = \|\mathbf{P}\|/\|\mathbf{J}\|$ выбран с таким расчетом, чтобы норма $c\mathbf{J}$ оставалась равной норме предельной матрицы \mathbf{P} . Используя правило, сформулированное в конце предыдущего раздела, можно для каждого t найти непрерывную компоненту спектра $[\lambda_1(t), \lambda_2(t)]$. Результат вычислений показан на рис.3 для $\alpha = 0.5$. Асимптотика $\lambda_{1,2}(t)$ имеет вид $\lambda_{1,2}(t) \simeq 1 \mp 2\sqrt{\alpha}/\sqrt{t}$, $t \rightarrow \infty$. Таким образом, в пределе больших t плотность спектра $(1 - \alpha)\delta(\lambda) + p(\lambda, t)$ аппроксимирует спектральную плотность проектор-матрицы $(1 - \alpha)\delta(\lambda) + \alpha\delta(\lambda - 1)$. "Ухудшение" матрицы связности в начальный период разобучения проявляется в некотором расширении непрерывной части спектра, как видно из рис.3.

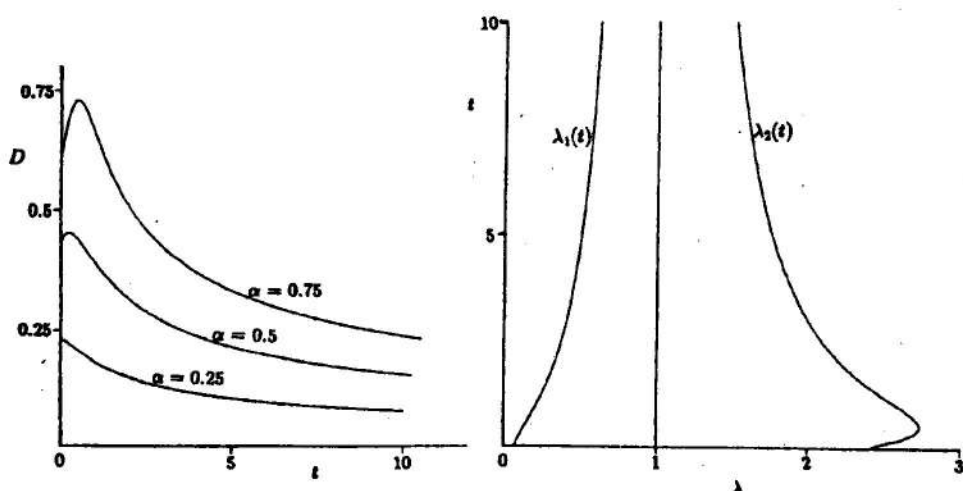


Рис. 2, 3.

4. ЗАКЛЮЧИТЕЛЬНЫЕ ЗАМЕЧАНИЯ

В настоящей статье продолжено изучение сходящегося алгоритма разобучения, предложенного в статьях [7, 8]. Выведены уравнения спектральной динамики для синаптической матрицы. Решена задача оптимизации алгоритма, для оптимального разобучения определены скорость сходимости к проектор-матрице, а также динамика спектра синаптической матрицы.

Представляет интерес исследование алгоритма с постоянной силой разобучения. Кроме того, интересно получить в явном виде выражение для спектральной плотности синаптической матрицы. Эти вопросы будут рассмотрены в следующей статье.

Автор выражает благодарность С.А.Семенову за многочисленные полезные обсуждения, а также Т.В.Вечерской и И.Б.Шуваловой за помощь в подготовке рукописи.

ЛИТЕРАТУРА

1. Hjjpfield J.J. // Proc.Natl.Acad.Sci.USA. 1982. V.79. N 8. P.2554.
2. Amit D.J., Gutfreund H., Sompolinsky H. // Ann.Phys. 1987. V.173. N 1. P.30.
3. Diederich S., Opper M. // Phys.Rev.Lett. 1987. V.58. N 9. P.949.
4. Krauth W., Mezard M. // J.Phys.A. 1987. V.20. N 11. L745.
5. Blatt M.G., Vergini E.G. // Phys.Rev.Lett. 1991. V.66.N 13. P.1793.
6. Berryman K.W., Inchiosa M.E., Jaffe A.M., Janowsky S.A. // J.Phys.A. 1990. V.23. N 5. L223.

7. Plakhov A. Yu., Semenov S. A. // *Neural Network World*. 1993. V. 4. P. 405.
8. Plakhov A. Yu., Semenov S. A. // *J. Phys. I (France)*. 1994. V. 4. N 2.
9. Personnaz L., Guyon I., Dreyfus G. // *J. Phys. (France) Lett.* 1985. V. 46. N 8. L359.
10. Kanter I., Sompolinsky H. // *Phys. Rev. A*. 1987. V. 35. N 1. P. 380.
11. Марченко В. А., Пастур Л. А. // *Матем сб.* 1967. Т. 72(114). N 4. С. 507.
12. Oppen M. // *Europhys. Lett.* 1989. V. 8. N 4. P. 389.

Институт физико-технических
проблем, Москва

Поступила в редакцию
11 декабря 1993 г.

THE UNLEARNING DYNAMICS IN THE HOPFIELD NEURAL NETWORK

A. Yu. Plakhov

The dynamics of the converging iterative unlearning algorithm proposed in refs. [7, 8] is examined in the paper. In the thermodynamical limit the self-consistent system of equations of spectral dynamics of synaptic matrix is obtained. The (changing in the course of iteration) unlearning strength which optimizes the convergence rate of algorithm to the projector matrix is found. For the optimal unlearning the dynamics of spectrum of synaptic matrix is solved.