Севастопольский государственный университет
Институт информационных технологий
Дополнительная профессиональная программа профессиональной переподготовки «Глубокие нейросети в компьютерном зрении»

Основы нейронных сетей

Лекция 4
Линейный адаптивный элемент и правило его обучения

Бондарев Владимир Николаевич

ЛИНЕЙНЫЕ ОДНОСЛОЙНЫЕ СЕТИ

План

1. Линейный адаптивный нейрон ADALINE
2. Классификация с использованием линейного нейрона
3. Критерий и решение минимума среднего квадрата ошибки (MSE)
4. LMS алгоритм и условия его сходимости

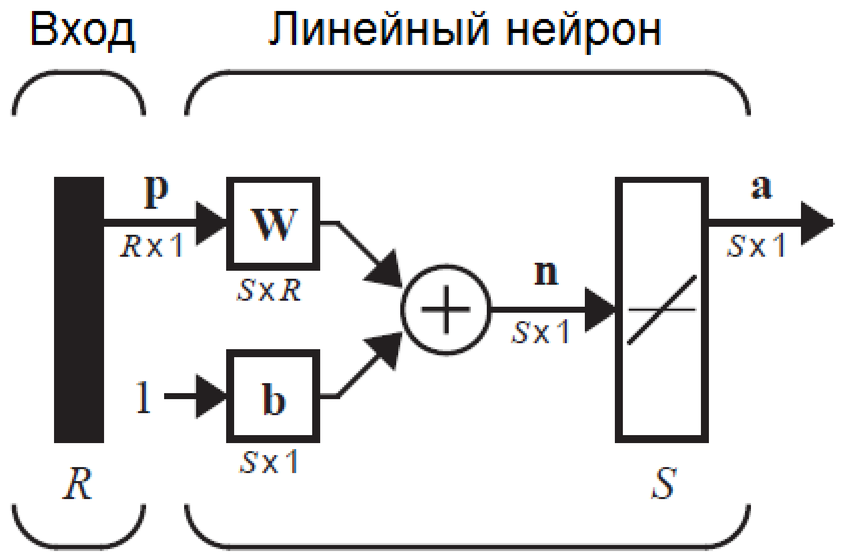
Линейный адаптивный нейрон - ADALINE

В 1960 г. Б.Уидроу и его студент М. Хофф предложили адаптивный линейный нейрон ADALINE и правило обучения, которое они назвали **LMS** (Least Mean Square**) алгоритмом**

Нейрон **ADALINE подобен персептрону**, кроме функции преобразования, которая является линейной.
ADALINE имеет те же недостатки: решает только линейные сепарабельные задачи.
Однако, **алгоритм LMS эффективнее**, чем правило обучения персептрона, т.к. удаляет границу решения, насколько это возможно, от образцов, используемых при обучении.

Алгоритм **LMS шире применяется на практике**, особенно в области обработки сигналов.

Однослойная сеть ADALINE



$$\mathbf{a} = \mathbf{purelin}(\mathbf{Wp}+\mathbf{b}) = \mathbf{Wp}+\mathbf{b}$$

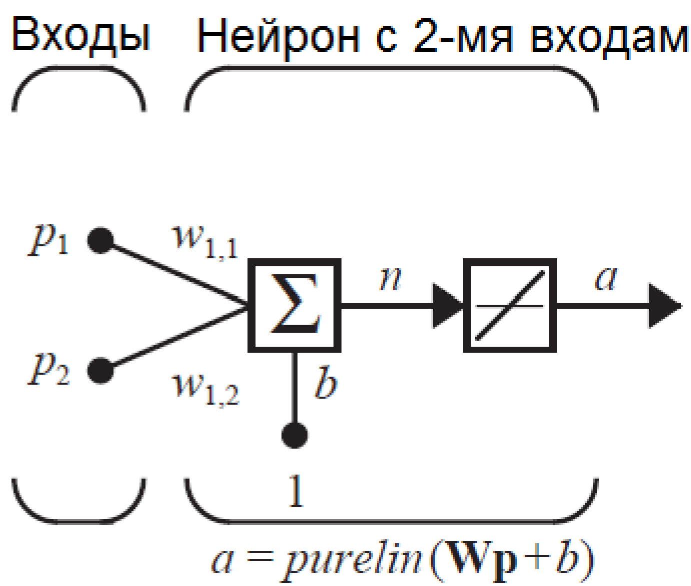
$$a_i = purelin(n_i) = purelin({}_i\mathbf{w}^T\mathbf{p} + b_i) = {}_i\mathbf{w}^T\mathbf{p} + b_i ,$$

где $\mathbf{w}_i$ строка матрицы **W** :

$${}_i\mathbf{w} = \begin{bmatrix} w_{i,1} \\ w_{i,2} \\ \vdots \\ w_{i,R} \end{bmatrix} .$$

Классификация с использованием линейного нейрона

Для упрощения рассуждений рассмотрим один нейрон с 2-мя входами и 1 выходом.



$$a = purelin(n) = purelin({}_1\mathbf{w}^T\mathbf{p} + b) = {}_1\mathbf{w}^T\mathbf{p} + b$$

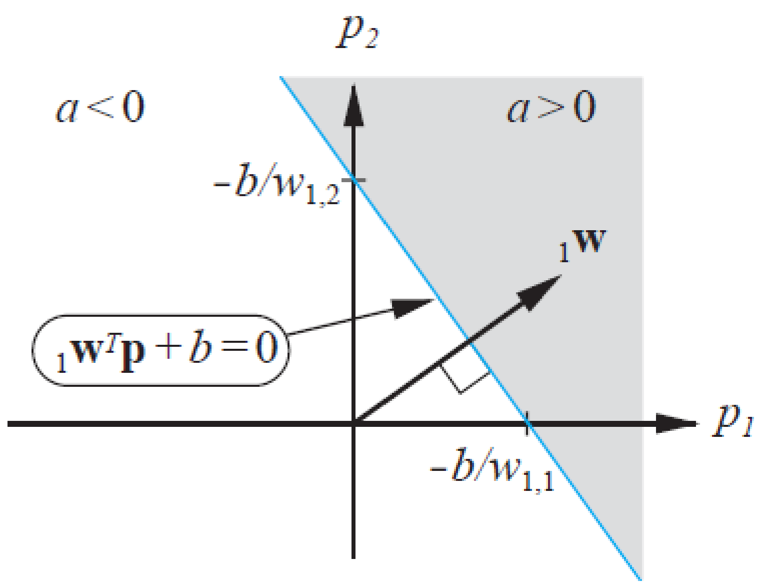
$$= {}_1\mathbf{w}^T\mathbf{p} + b = w_{1,1}p_1 + w_{1,2}p_2 + b \, .$$

Граница решения ADALINE

Аналогично персептрону, граница решения ADALINE определяется условием n=a=0 . Решив уравнение

$$\mathbf{w}^T\mathbf{p} + b = 0 \quad ,$$

можно построить границу.



Т.о. ADALINE (как и персептрон) может использоваться для разделения объектов на 2 категории, но только **линейно сепарабельных.**

Критерий среднего квадрата ошибки (MSE)

Рассмотрим обучение с учителем. Пусть обучающее множество содержит входные образы $\mathbf{p}_q$ и желаемые выходы $\mathbf{t}_q$:

$$\{\mathbf{p}_1, \mathbf{t}_1\}, \{\mathbf{p}_2, \mathbf{t}_2\}, \ldots, \{\mathbf{p}_Q, \mathbf{t}_Q\}$$

Для упрощения анализа введем следующие обозначения:

$$\mathbf{x} = \begin{bmatrix} {}_1\mathbf{w} \\ b \end{bmatrix} \qquad \mathbf{z} = \begin{bmatrix} \mathbf{p} \\ 1 \end{bmatrix} \qquad a = \mathbf{x}^T\mathbf{z}.$$

Целевая функция - средний квадрат ошибки (MSE)

$$F(\mathbf{x}) = E[e^2] = E[(t-a)^2] = E[(t-\mathbf{x}^T\mathbf{z})^2],$$

где E – символ математического ожидания , запишется в форме :

$$F(\mathbf{x}) = E[t^2 - 2t\mathbf{x}^T\mathbf{z} + \mathbf{x}^T\mathbf{z}\mathbf{z}^T\mathbf{x}] = E[t^2] - 2\mathbf{x}^T E[t\mathbf{z}] + \mathbf{x}^T E[\mathbf{z}\mathbf{z}^T]\mathbf{x}.$$

Отсюда MSE целевая функция может быть переписана в квадратичной форме

$$F(\mathbf{x}) = c - 2\mathbf{x}^T\mathbf{h} + \mathbf{x}^T\mathbf{R}\mathbf{x}, \text{ где } c = E[t^2], \mathbf{h} = E[t\mathbf{z}] \text{ and } \mathbf{R} = E[\mathbf{z}\mathbf{z}^T]. \qquad (1)$$

Здесь **h** – вектор кросс-корреляции; **R** – корреляционная матрица.

Решение минимума среднего квадрата ошибки (MSE)

Сравним MSE целевую фvнкцию

$$F(\mathbf{x}) = c - 2\mathbf{x}^T\mathbf{h} + \mathbf{x}^T\mathbf{R}\mathbf{x} \tag{1}$$

с квадратичной функцией

$$F(\mathbf{x}) = c + \mathbf{d}^T\mathbf{x} + \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x}.$$

Тогда $\mathbf{d} = -2\mathbf{h}$ и $\mathbf{A} = 2\mathbf{R}$.

В этом случае матрице Гессе соответствует **2R**. Корреляционная матрица всегда п.о. или п.п.о. Поэтому целевая функция (1) будет иметь один глобальный минимум, слабый минимум или минимумов не будет, в зависимости от $\mathbf{d} = -2\mathbf{h}$.

Найдем стационарную точку MSE целевой функции $F(\mathbf{x})$:

$$\nabla F(\mathbf{x}) = \nabla\left(c + \mathbf{d}^T\mathbf{x} + \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x}\right) = \mathbf{d} + \mathbf{A}\mathbf{x} = -2\mathbf{h} + 2\mathbf{R}\mathbf{x}.$$

Если корреляционная матрица п.о., то будет существовать единственная стационарная точка со строгим минимумом

$$\mathbf{x}^* = \mathbf{R}^{-1}\mathbf{h}.$$

Будем называть это решение – **решением минимума MSE.**

К сожалению, **существование строгого минимума целиком зависит от R**. Следовательно, характеристики входного вектора определяют существование единственного решения.

LMS алгоритм (наименьших средних квадратов)

На практике вычислять оценки обратной корреляционной матрицы **R** и вектора кросс-корреляции **h** неудобно.

Поэтому будем использовать аппроксимацию алгоритма SDA, где будут использоваться оценки градиента. Идея алгоритма LMS заключается в аппроксимации среднего квадрата ошибки $F(\mathbf{x})$ с помощью

$$\hat{F}(\mathbf{x}) = (t(k) - a(k))^2 = e^2(k),$$

где **мат.ожидание квадрата ошибки заменено на квадрат ошибки**.

Тогда оценка градиента, называемая **стохастическим градиентом** и вычисляемая на каждом шаге, будет равна

$$\hat{\nabla} F(\mathbf{x}) = \nabla e^2(k).$$

Т.к. $a = \mathbf{x}^T \mathbf{z}$, то $\hat{\nabla} F(\mathbf{x}) = \nabla e^2(k) = -2e(k)\mathbf{z}(k)$.

Подставим в алгоритм SDA

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha \nabla F(\mathbf{x})\Big|_{\mathbf{x}=\mathbf{x}_k}.$$

стохастический градиент, получим **LMS алгоритм (правило обучения WH)**

$$\mathbf{x}_{k+1} = \mathbf{x}_k + 2\alpha e(k)\mathbf{z}(k), \qquad \mathbf{x} = \begin{bmatrix} {}_1\mathbf{w} \\ b \end{bmatrix} \qquad \mathbf{z} = \begin{bmatrix} \mathbf{p} \\ 1 \end{bmatrix}$$

или $\quad {}_1\mathbf{w}(k+1) = {}_1\mathbf{w}(k) + 2\alpha e(k)\mathbf{p}(k), \quad b(k+1) = b(k) + 2\alpha e(k).$

$$\mathbf{W}(k+1) = \mathbf{W}(k) + 2\alpha \mathbf{e}(k)\mathbf{p}^T(k), \quad \mathbf{b}(k+1) = \mathbf{b}(k) + 2\alpha \mathbf{e}(k).$$

Сходимость алгоритма LMS

Сходимость SDA определялась значением собственных чисел матрицы Гессе A

$$\alpha < \frac{2}{\lambda_i}.$$

Для LMS справедливо, что $\mathbf{A} = 2\mathbf{R}$. Поэтому по аналогии можно записать условие сходимости LMS алгоритма

$$\alpha < 1/\lambda_i \text{ для всех } i \quad ,$$

или

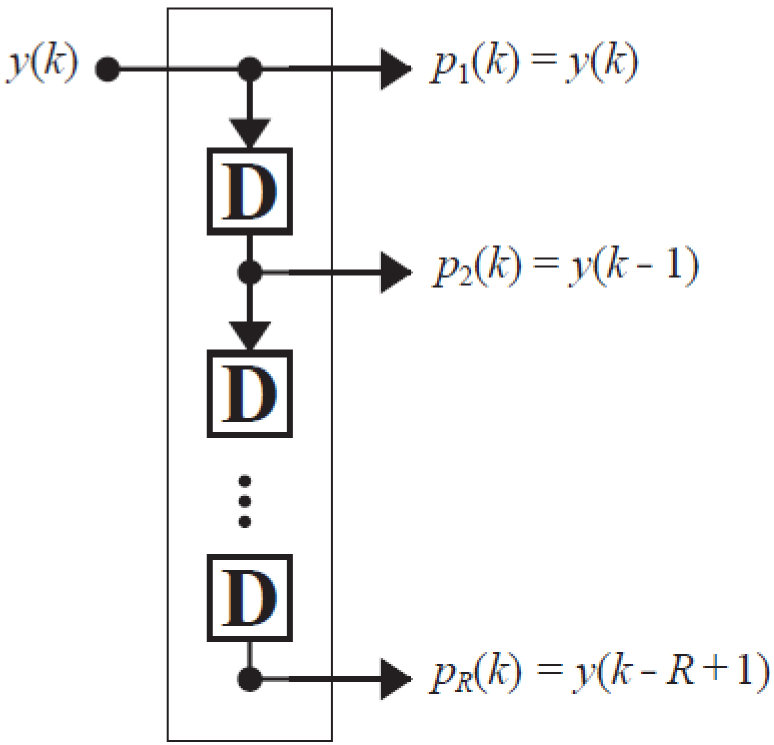
$$0 < \alpha < 1/\lambda_{max}.$$

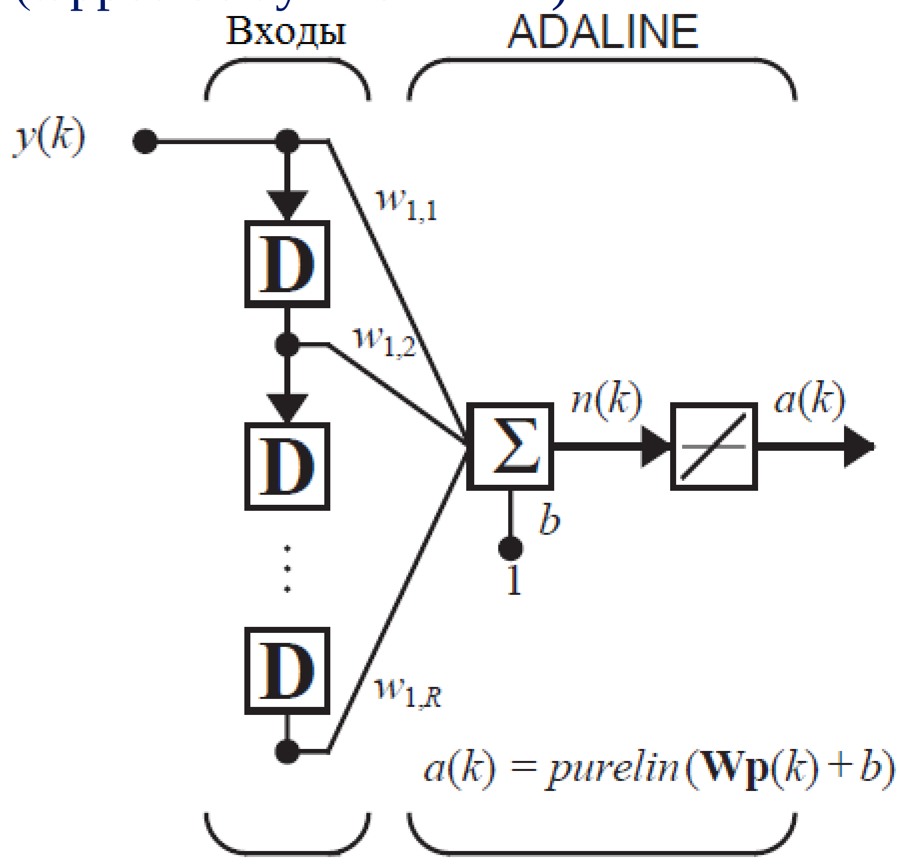
где λ_i - собственные числа корреляционной матрицы **R** (λ_i >0) . Если это условие выполняется, то можно показать, что LMS алгоритм, обрабатывающий на каждой итерации один входной вектор, сходится к решению, совпадающему с решением минимума среднего квадрата ошибки (MSE):

$$\mathbf{x}^* = \mathbf{R}^{-1}\mathbf{h}.$$

Адаптивная фильтрация

На основе сети ADALINE можно реализовать адаптивный фильтр. Для этого в её состав вводят линию задержки (tapped delay line – TDL)



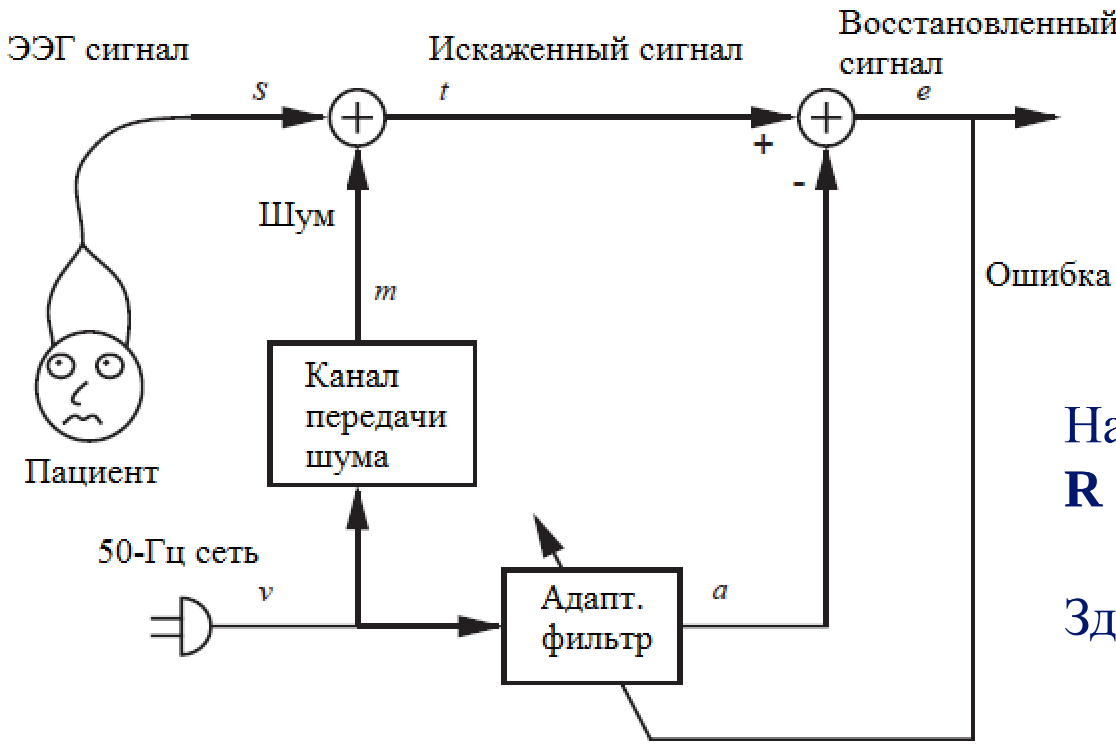


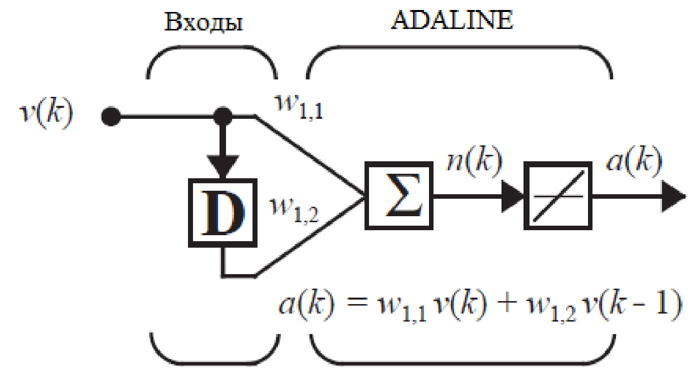
Выход адаптивного фильтра КИХ (FIR) фильтра:

$$a(k) = purelin(\mathbf{Wp} + b) = \sum_{i=1}^{R} w_{1,i}\, y(k - i + 1) + b\,.$$

Адаптивная фильтрация

Пример. Адаптивное подавление шума ЭЭГ. Т.к. шум является гармоническим сигналом, то для решения задачи можно применить АФ с 2-мя коэффициентами.





Найдем корреляционную матрицу **R** и вектор кросс-корреляции **h**

$$\mathbf{R} = [\mathbf{z}\mathbf{z}^T], \quad \mathbf{h} = E[t\mathbf{z}].$$

Здесь

$$\mathbf{z}(k) = \begin{bmatrix} v(k) \\ v(k-1) \end{bmatrix}, \quad t(k) = s(k) + m(k).$$

$$\mathbf{R} = \begin{bmatrix} E[v^2(k)] & E[v(k)v(k-1)] \\ E[v(k-1)v(k)] & E[v^2(k-1)] \end{bmatrix}, \quad \mathbf{h} = \begin{bmatrix} E[(s(k)+m(k))v(k)] \\ E[(s(k)+m(k))v(k-1)] \end{bmatrix}.$$

Адаптивная фильтрация

Пусть ЭЭГ сигнал – белый шум с равномерным распределением значений от -0.2 до +0.2. Шум – синусоид. сигнал с частотой 50 Гц и частотой отсчетов 150 Гц

$$v(k) = 1.2 \sin\left(\frac{2\pi k}{3}\right),$$

а шум, искажающий ЭЭГ сигнал, – его ослабленная и смещенная по фазе копия

$$m(k) = 0.12 \sin\left(\frac{2\pi k}{3} + \frac{\pi}{2}\right).$$

В этом случае (при вычислении среднего на периоде синусоиды, k=1..3)

$$\mathbf{R} = \begin{bmatrix} 0.72 & -0.36 \\ -0.36 & 0.72 \end{bmatrix}. \qquad \mathbf{h} = \begin{bmatrix} 0 \\ -0.0624 \end{bmatrix}.$$

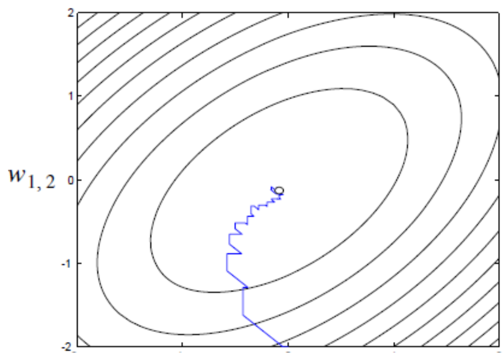
MSE решение

$$\mathbf{x}^* = \mathbf{R}^{-1}\mathbf{h} = \begin{bmatrix} 0.72 & -0.36 \\ -0.36 & 0.72 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ -0.0624 \end{bmatrix} = \begin{bmatrix} -0.0578 \\ -0.1156 \end{bmatrix}.$$

Можно показать, что $E[s^2(k)]=0.0133$ и ЦФ будет тоже равна $F(\mathbf{x}^*)=0.0133$. Т.е. минимум СКО равен среднему квадрату ЭЭГ сигнала и, следовательно, сигнал ошибки на выходе схемы является восстановленным ЭЭГ сигналом.

Адаптивная фильтрация

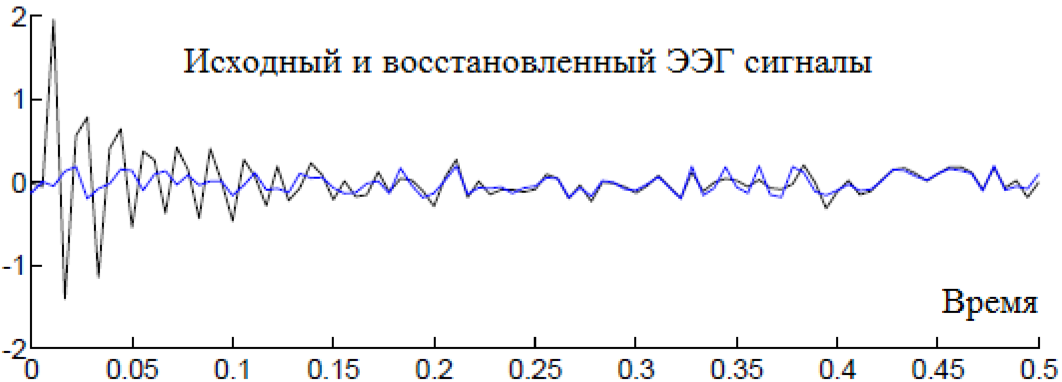
Если при решении задачи применить LMS алгоритм, то при α=0.1 получим траекторию спуска, подобную зашумленной версии наискорейшего спуска

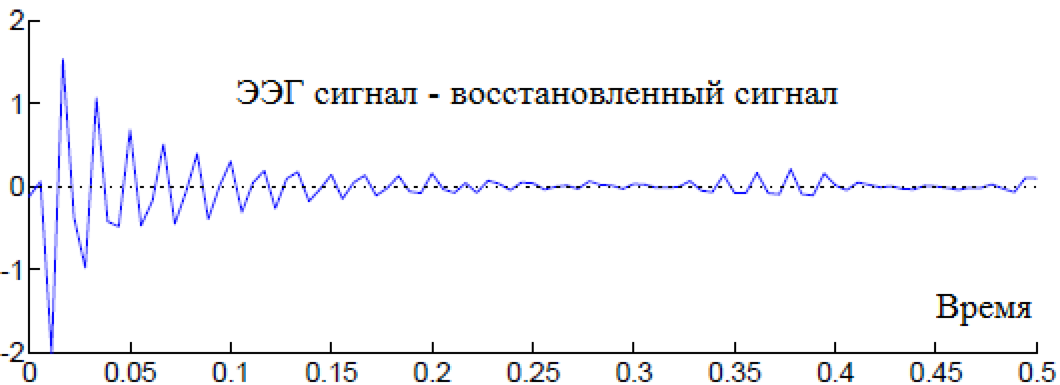


Собственные значения и собственные вектора Гессиана **A=2R**:

$$\lambda_1 = 2.16\,,\ \mathbf{z}_1 = \begin{bmatrix} -0.7071 \\ 0.7071 \end{bmatrix},\ \lambda_2 = 0.72\,,\ \mathbf{z}_2 = \begin{bmatrix} -0.7071 \\ -0.7071 \end{bmatrix}.$$

Максимально возможное устойчивое значение α<2/2.16=0.926





Отличие сигналов не равно нулю, т.к. LMS алгоритм является аппроксимацией наискорейшего спуска; он использует оценку градиента, а не его точное значение при обновлении весов. Это приводит к тому, что веса продолжают изменяться даже после достижения точки минимума СКО.