

## Теория планирование проведения эксперимента

### Лекция 1

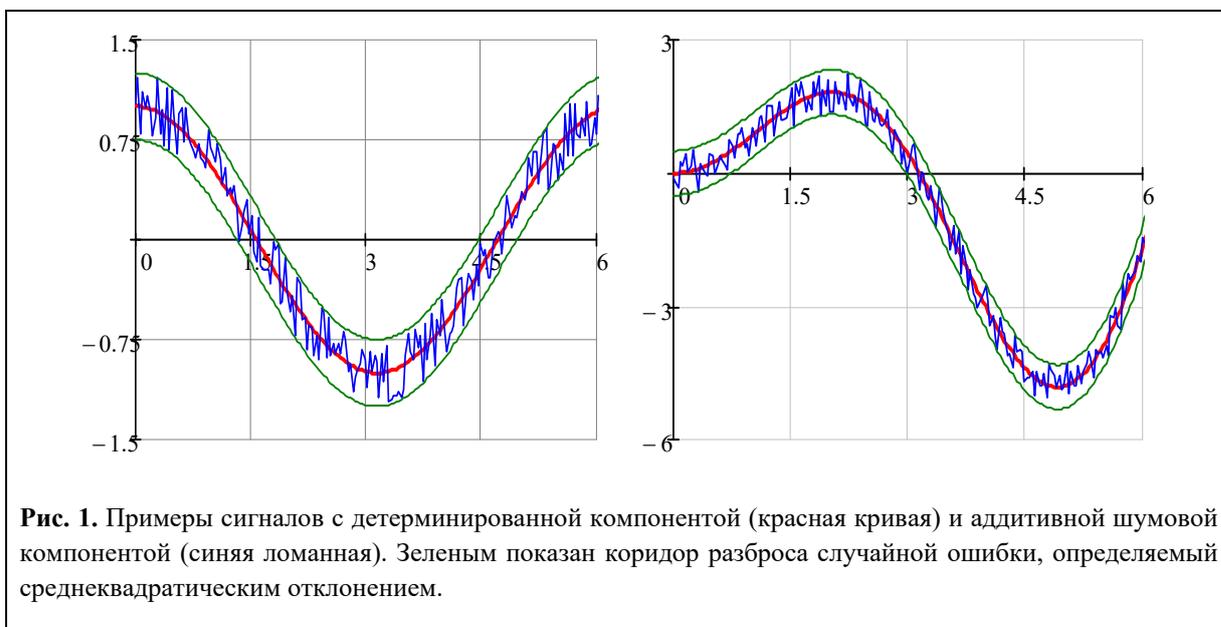
#### Измерение сигналов и сглаживание шумов (скользящее среднее).

При экспериментальном исследовании, какого либо технологического процесса, измеряемый результат последнего является случайной величиной, на которую оказывают влияния много факторов. Результат исследования является, как правило, случайной величиной распределенной по какому ни будь закону. Чаще всего по нормальному закону. Будем называть измеряемые величины сигналами.

Измеряемые сигналы всегда состоят из суммы детерминированной и случайной составляющих. То есть происходит аддитивное наложение шума (addition- сложение) на полезный сигнал.

$$y(t) = y_k + \Delta_k \quad (1)$$

Здесь  $y_k$  - детерминированная часть сигнала (полезный сигнал),  $\Delta_k$  - шумовая составляющая сигнала.



**Рис. 1.** Примеры сигналов с детерминированной компонентой (красная кривая) и аддитивной шумовой компонентой (синяя ломаная). Зеленым показан коридор разброса случайной ошибки, определяемый среднеквадратическим отклонением.

Природа шума может быть разной. Например, это может быть тепловые помехи (сигнал транзистора), это может быть конечный разряд цифровой системы в случае дискретного сигнала с АЦП, это может быть ограниченная точность прибора, и т. д.

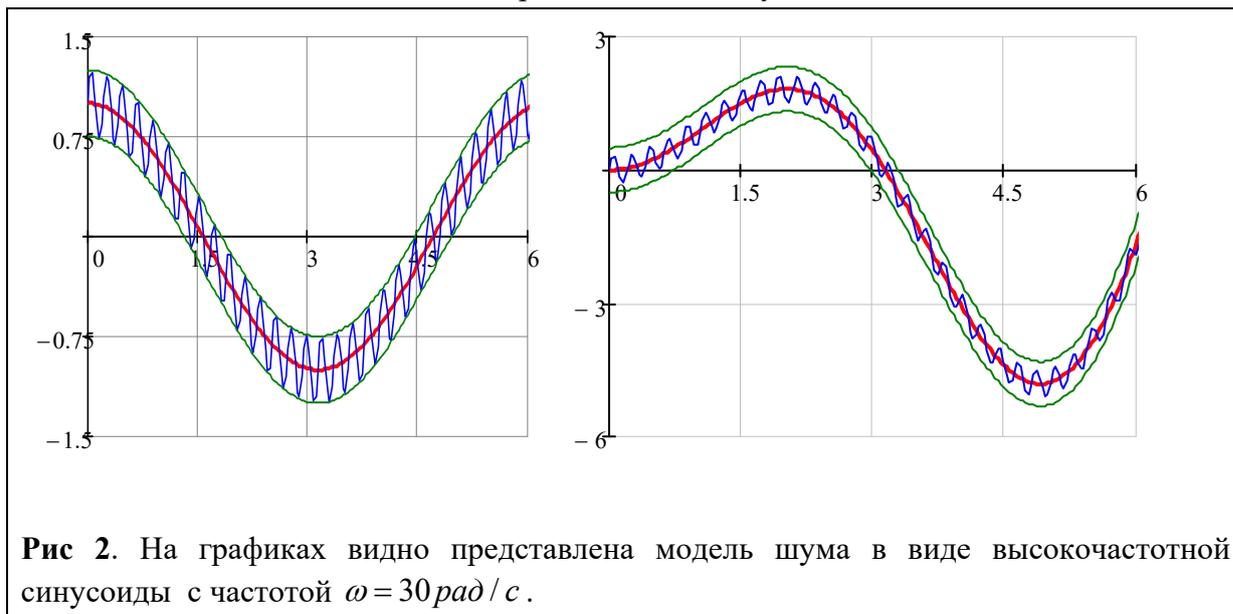
Таким образом, сигнал имеет ошибку. Различают ошибки абсолютные и относительные. Абсолютная ошибка – это отклонение сигнала от его точного значения.

$$\Delta = y - y_k \quad (2)$$

Относительная ошибка это отношение абсолютной ошибки к величине сигнала в процентах

$$\frac{\Delta}{y} 100 = \frac{y - y_k}{y} 100.$$

Если посмотреть на рисунок 1, то можно увидеть, что шумовая составляющая имеет высокие частоты и её можно моделировать в виде синусоиды с высокой частотой.



$$y(t) = y_1(t) + \Delta \sin(\omega t), \quad \omega = 30 \text{ рад} / \text{с}$$

Продифференцируем сигнал

$$y'(t) = y_1'(t) + \omega \Delta \cos(\omega t)$$

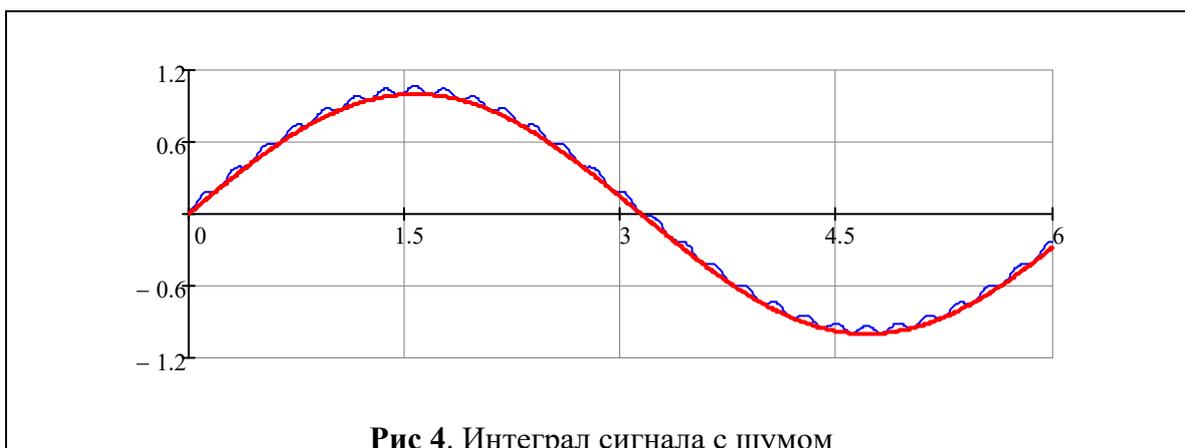


*При дифференцировании относительный вклад шумовой составляющей увеличивается, а относительный вклад полезного сигнала ослабляется.* Поэтому операцию дифференцирования следует делать аккуратно. Её вообще избегают.

Теперь рассмотрим, что происходит с сигналом при интегрировании. Снова выберем модель шума в виде синусоиды

$$y(t) = y_1(t) + \Delta \sin(\omega t), \quad \omega = 30 \text{ рад} / \text{с}$$

$$\int y(t) dt = \int y_1(t) dt + \frac{\Delta \cos(\omega t)}{\omega}$$



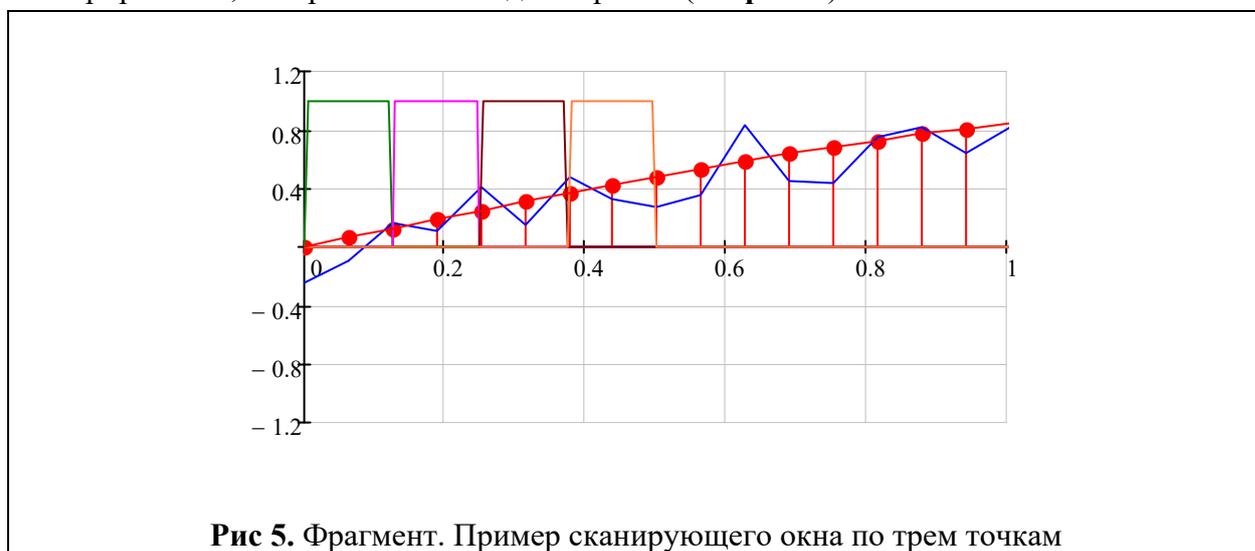
**Рис 4.** Интеграл сигнала с шумом

На рисунке 4 показано, что относительный вклад шумовой составляющей сигнала существенно уменьшился. Таким образом, при интегрировании относительный вклад шумовой составляющей уменьшается. Если использовать интегрирование на малом диапазоне, можно без существенного искажения полезного сигнала уменьшать относительный вклад шумовой составляющей. Такой метод называется скользящее среднее.

### Метод скользящего среднего

Выберем совокупность дискретных значений сигнала. Такую совокупность принято называть выборками или уровнями (в иностранной литературе time series). При первом взгляде на график можно увидеть тенденцию изменения измеряемой величины, так называемый тренд, и его колебательную составляющую. Тренд это полезная составляющая сигнала. Иногда колебательная составляющая вносит значительный вклад и на фоне шума трудно увидеть тренд. На практике для обнаружения тренда используют метод скользящего среднего.

Метод скользящего среднего заключается в том, что выбирается некоторое окно интегрирования, которое скользит вдоль кривой (см. **рис 5.**)



**Рис 5.** Фрагмент. Пример сканирующего окна по трем точкам

В пределах окна происходит интегрирование (численное) или просто сложение точек в пределах окна и усреднение по точкам этого окна.

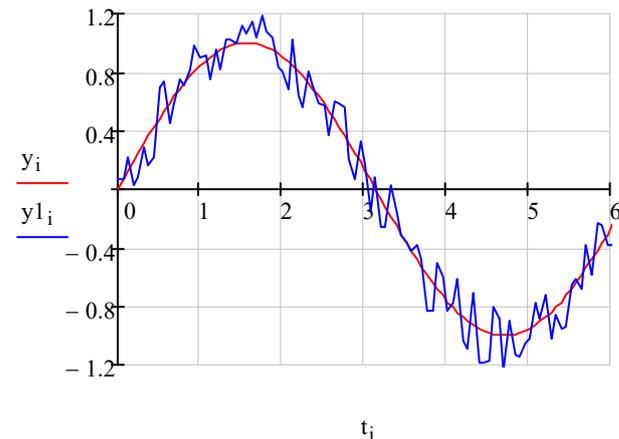
Приведем алгоритм в программе *Mathcad*

$N := 100$  число выборок  $i := 0..N$   $t_i := \frac{2 \cdot \pi}{N} \cdot i$  интервал делится на дискретное множество

$t_i$	0	1	2	3	4	5	6	7	8
	0	0.063	0.126	0.188	0.251	0.314	0.377	0.44	...

$\sigma := 0.5$   $n_i := \sigma \cdot (\text{rnd}(1) - 0.5)$  моделируется наброс шума с равномерным распределением

$y_i := \sin(t_i)$   $y1_i := y_i + n_i$  . Строится график с шумом.



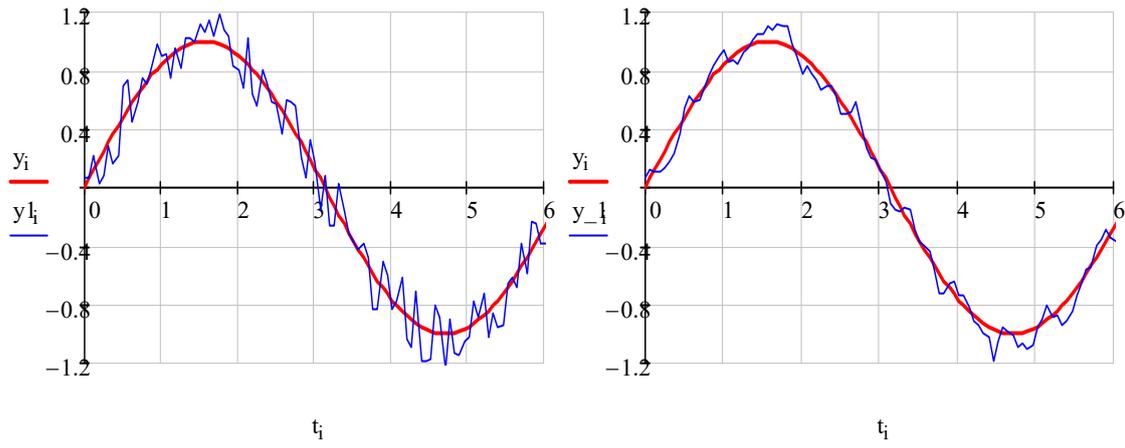
Выборки, уровни

$y =$	0	0	$y1 =$	0	0.064
	1	0.063		1	0.065
	2	0.125		2	0.223
	3	0.187		3	0.032
	4	0.249		4	0.088
	5	0.309		5	0.288
	6	0.368		6	0.167
	7	0.426		7	0.223
	8	0.482		8	0.697
	9	0.536		9	0.733
	10	0.588		10	0.451
	11	0.637		11	0.593
	12	0.685		12	0.749
	13	0.729		13	0.705
	14	0.771		14	0.819
	15	...		15	...

Произведем два усреднения усреднение в пределах окна (по трем точкам)

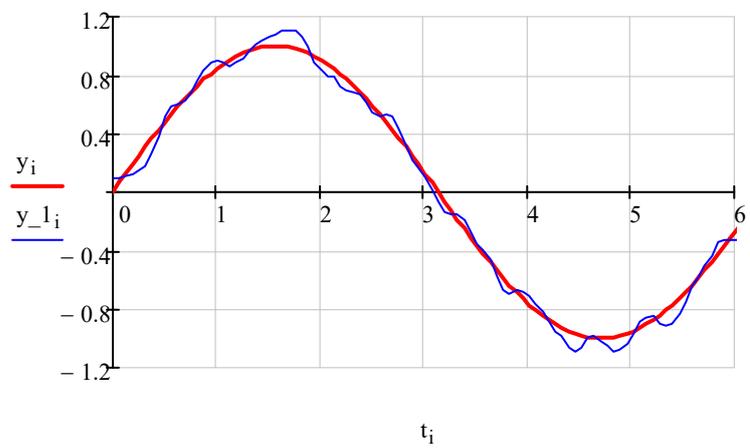
$$i := 0..N - 1 \quad y_{-1}_i := \text{if} \left( i - 1 < 0, \frac{0 + y1_i + y1_{i+1}}{2}, \frac{y1_{i-1} + y1_i + y1_{i+1}}{3} \right)$$

Для сравнения рисуем график без усреднения и с усреднением.



Сглаживания оказалось не достаточно, потому сглаживаем еще раз.

$$i := 0..N - 1 \quad y1_i := y_{-1_i} \quad y_{-1_i} := \text{if} \left( i - 1 < 0, \frac{0 + y1_i + y1_{i+1}}{2}, \frac{y1_{i-1} + y1_i + y1_{i+1}}{3} \right)$$



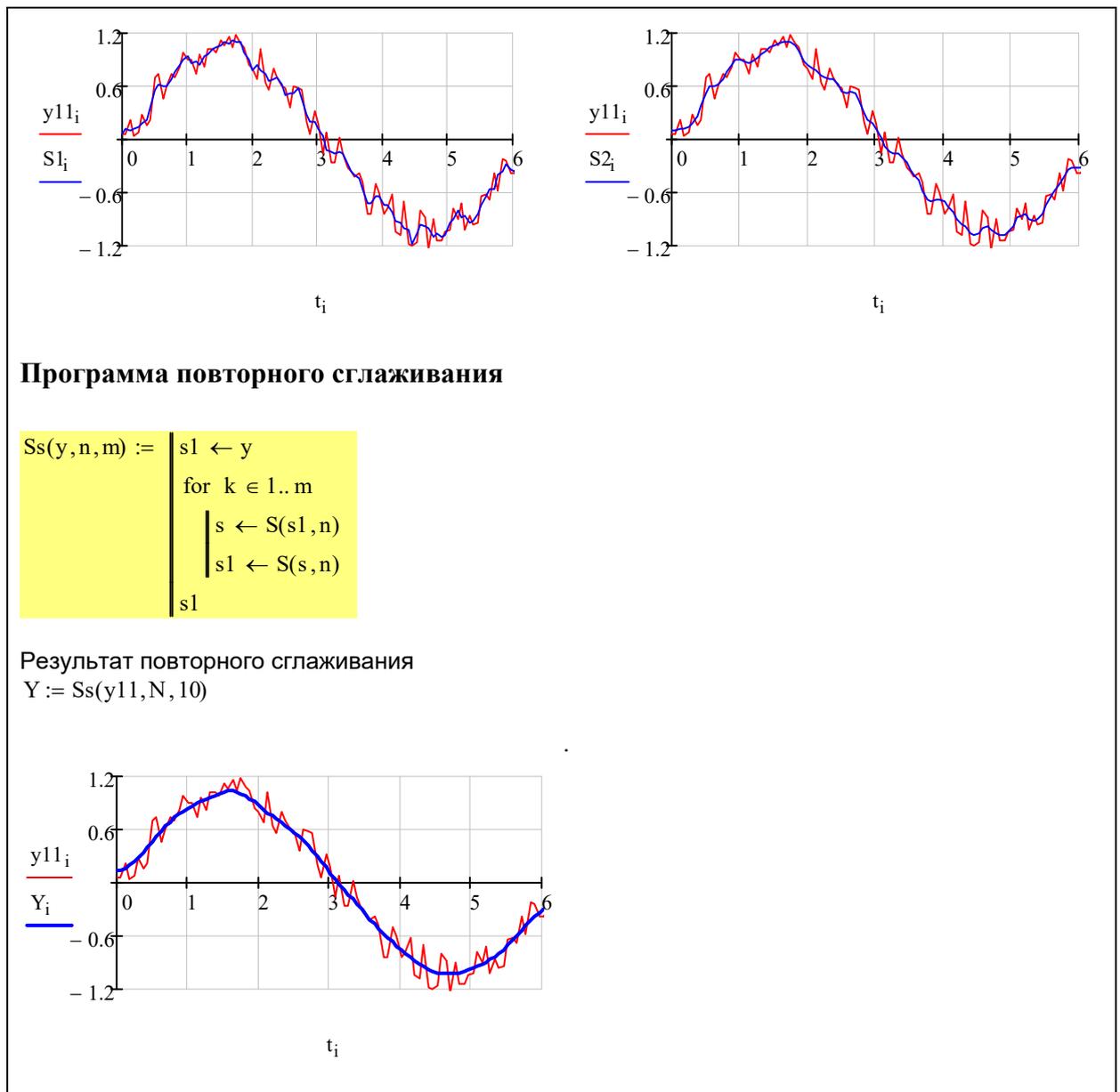
Затем процедура повторяется желаемое количество раз. Для этого нужно писать программу

**Программа сглаживания по трём точкам**

```

S(y,n) :=
  y1_0 ← (y_0 + y_1) / 2
  y1_n ← (y_n + y_{n-1}) / 2
  for i ∈ 1..n - 1
    y1_i ← (y_{i-1} + y_i + y_{i+1}) / 3
  y1
  
```

S1 := S(y11,N)    S2 := S(S1,N)



В общем случае метод скользящего среднего опирается на *теорему Вейерштрасса*, согласно которой любая гладкая функция при самых общих допущениях может быть локально (то есть в *ограниченном интервале измерений*) представлена алгебраическим *полиномом подходящей степени*. То есть идея заключается в том, что бы полезную, детерминированную составляющую локально, на активном участке представлять в виде полином не высокой степени. Полином может быть первой, второй или третьей степени.

Выбирается активный участок с нечетным количеством точек и на этом интервале производится приближение функции полиномом первой степени  $y_k = a_0 + a_1k$ , с дальнейшим использованием метода наименьших квадратов.

Выберем активный участок - интервал сглаживания из  $l = 2n + 1$  точек. И запишем приближение функции на этом участке. Для удобства расчета перенесем начало координат в центр активного участка как показано на **рис 6**. Коэффициенты линейного

приближения на активном участке рассчитываются методом наименьших квадратов (МНК).

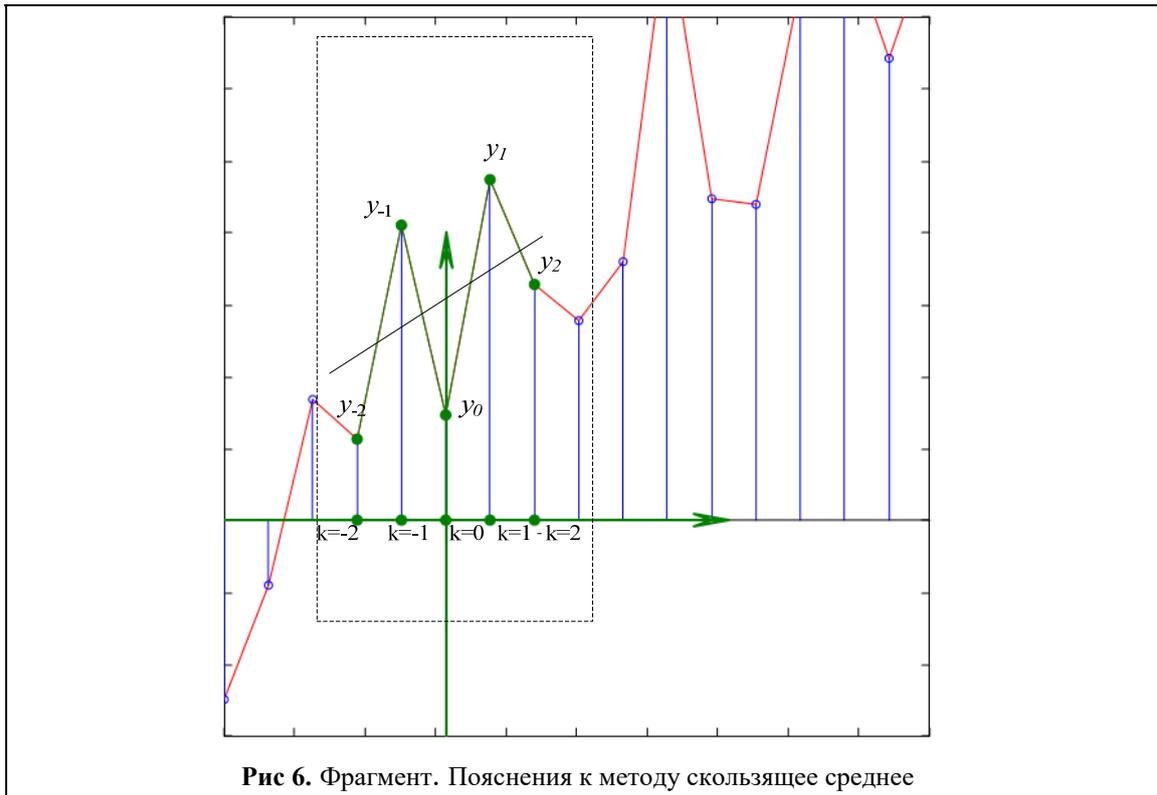


Рис 6. Фрагмент. Пояснения к методу скользящее среднее

Запишем необходимые уравнения для определения коэффициентов МНК. Для этого запишем целевую функцию и найдем её минимум, то есть найдем минимум отклонения заданной функции от её приближения

$$Q = \sum_{k=-n}^{k=n} (y_k - a_0 - a_1 k)^2 \rightarrow \min$$

В этом выражении нам неизвестны коэффициенты  $a_0, a_1$ . Что бы найти коэффициенты необходимо взять производные целевой функции  $Q$  по коэффициентам и приравнять нулю.

$$\frac{\partial Q}{\partial a_0} = 2 \sum_{k=-n}^{k=n} (y_k - a_0 - a_1 k) = 0 \rightarrow \sum_{k=-n}^{k=n} y_k = a_0 (2n+1) + a_1 \sum_{k=-n}^{k=n} k$$

$$\frac{\partial Q}{\partial a_1} = 2 \sum_{k=-n}^{k=n} (y_k - a_0 - a_1 k) k = 0 \rightarrow \sum_{k=-n}^{k=n} y_k k = a_0 \sum_{k=-n}^{k=n} k + a_1 \sum_{k=-n}^{k=n} k^2$$

Из последней системы уравнений нам нужно найти коэффициент  $a_0$ . Он легко определяется. Учитывая что

$$\sum_{k=-n}^{k=n} k = 0$$

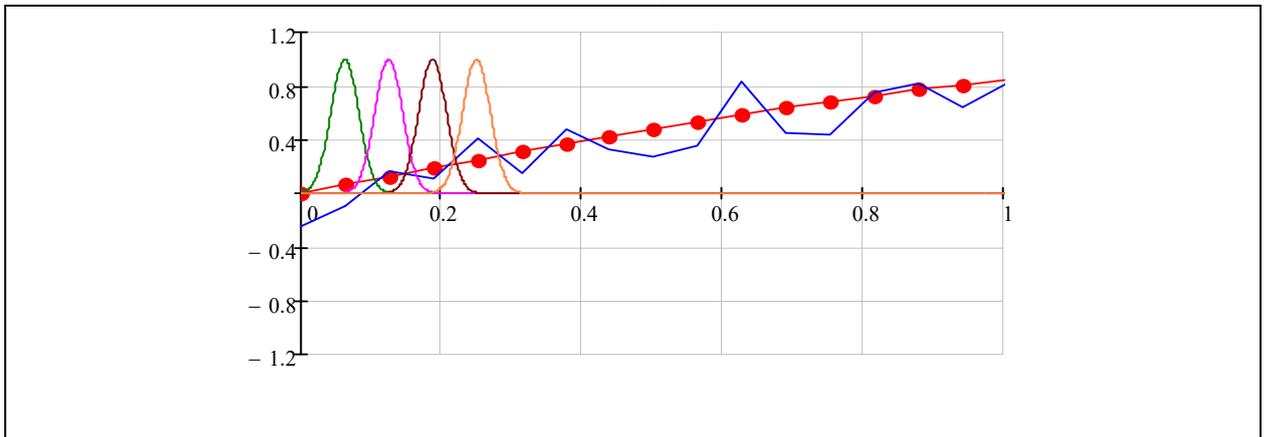
получаем

$$\sum_{k=-n}^{k=n} y_k = a_0 (2n+1) \rightarrow a_0 = \frac{\sum_{k=-n}^k y_k}{(2n+1)}, \hat{y}_0 = a_0$$

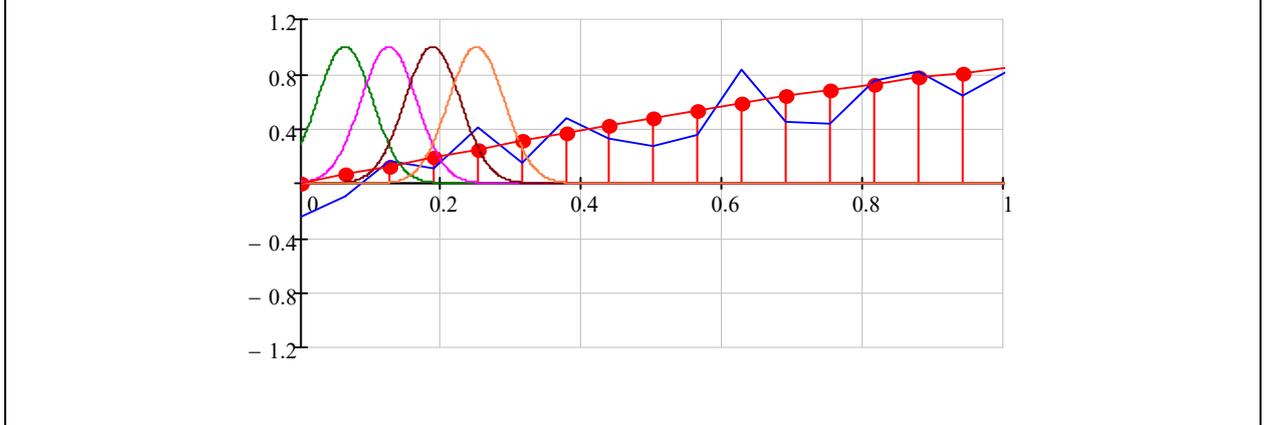
Таким образом, для того чтобы найти сглаженное значение функции в точке необходимо использовать её среднее значение на активном интервале.

Лучший результат получается при формировании весовых окон, потому что весовые окна учитывают относительный вклад каждой точки. *Относительный вклад* – значит, учитывает весовые коэффициенты точек, подчёркивая значимость одних точек перед другими. Очевидно, что основной вклад должна давать та точка, которая находится в середине – активная точка (точка в которой происходит измерение в данный момент), потому что её удельный вес в процессе более значим, чем остальных точек, удаленных от нее. В таком алгоритме проявляется некоторая оптимальность и корреляция. Интуитивно ясно, что чем удалённей точка, тем менее значим её вклад.

На рисунке **Рис 7**. Приведен пример оптимального сканирующего окна по трем точкам. Точек может быть и больше, тогда процесс сглаживания происходит быстрее, потому что происходит большее поглощение колебаний (см рис 8).



**Рис 7.** Пример оптимального сканирующего окна по трем точкам

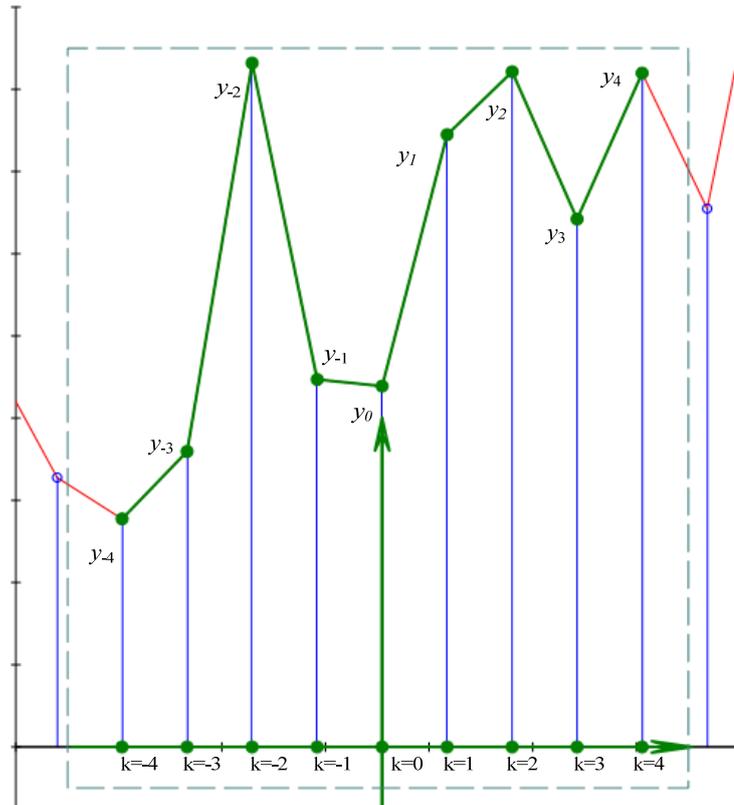


**Рис 8.** Пример оптимального сканирующего окна по пяти точкам.

### Использование взвешенных коэффициентов скользящего среднего

Мы заметили, что при использовании скользящего среднего все значения функции в активном интервале одинаковы. Если вести коэффициенты с весами, то значение функции в центральной точке будет определяться выражением:

$$\hat{y} = \frac{\sum_{k=-n}^{k=n} y_k W_k}{\sum_{k=-n}^{k=n} W_k},$$



При весовом методе скользящего среднего обычно используют полином второго или третьего порядка. Приведём пример. Выберем активный интервал с девятью точками  $l = 2n + 1$ ,  $n=3$ ,  $l=7$  и сместим ось в центр этого интервала. Запишем кубический полином

$$y_k = a_0 + a_1 k + a_2 k^2 + a_3 k^3$$

Далее записывают целевую функцию для определения коэффициентов  $a_0, a_1, a_2, a_3$

$$Q = \sum_{k=-n}^{k=n} (y_k - a_0 - a_1 k - a_2 k^2 - a_3 k^3)^2 \rightarrow \min$$

$$\frac{\partial Q}{\partial a_0} = 2 \sum_{k=-n}^{k=n} (y_k - a_0 - a_1 k - a_2 k^2 - a_3 k^3) = 0 \rightarrow \sum_{k=-n}^{k=n} y_k = a_0 (2n+1) + a_1 \sum_{k=-n}^{k=n} k + a_2 \sum_{k=-n}^{k=n} k^2 + a_3 \sum_{k=-n}^{k=n} k^3$$

$$\frac{\partial Q}{\partial a_1} = 2 \sum_{k=-n}^{k=n} (y_k - a_0 - a_1 k - a_2 k^2 - a_3 k^3) k = 0 \rightarrow \sum_{k=-n}^{k=n} y_k k = a_0 \sum_{k=-n}^{k=n} k + a_1 \sum_{k=-n}^{k=n} k^2 + a_2 \sum_{k=-n}^{k=n} k^3 + a_3 \sum_{k=-n}^{k=n} k^4$$

$$\frac{\partial Q}{\partial a_2} = 2 \sum_{k=-n}^{k=n} (y_k - a_0 - a_1 k - a_2 k^2 - a_3 k^3) k^2 = 0 \rightarrow \sum_{k=-n}^{k=n} y_k k^2 = a_0 \sum_{k=-n}^{k=n} k^2 + a_1 \sum_{k=-n}^{k=n} k^3 + a_2 \sum_{k=-n}^{k=n} k^4 + a_3 \sum_{k=-n}^{k=n} k^4$$

$$\frac{\partial Q}{\partial a_3} = 2 \sum_{k=-n}^{k=n} (y_k - a_0 - a_1 k - a_2 k^2 - a_3 k^3) k^3 = 0 \rightarrow \sum_{k=-n}^{k=n} y_k k^3 = a_0 \sum_{k=-n}^{k=n} k^3 + a_1 \sum_{k=-n}^{k=n} k^4 + a_2 \sum_{k=-n}^{k=n} k^5 + a_3 \sum_{k=-n}^{k=n} k^6$$

Подставим в место  $n$  число 4. В приведенной системе уравнений пропадают слагаемые с нечетными степенями

$$\sum_{k=-3}^{k=3} k = 0, \quad \sum_{k=-3}^{k=3} k^3 = 0, \quad \sum_{k=-3}^{k=3} k^5 = 0.$$

Тогда системы уравнений выглядит проще

$$\begin{cases} \sum_{k=-3}^{k=3} y_k = a_0 \cdot 7 + a_2 \sum_{k=3}^{k=3} k^2 \\ \sum_{k=-3}^{k=3} y_k k = a_1 \sum_{k=3}^{k=3} k^2 + a_3 \sum_{k=3}^{k=3} k^4 \\ \sum_{k=-3}^{k=3} y_k k^2 = a_0 \sum_{k=3}^{k=3} k^2 + a_2 \sum_{k=3}^{k=3} k^4 \\ \sum_{k=-3}^{k=3} y_k k^3 = a_1 \sum_{k=3}^{k=3} k^4 + a_3 \sum_{k=3}^{k=3} k^6 \end{cases} \rightarrow \begin{cases} \sum_{k=-3}^{k=3} y_k = 9a_0 + 28a_2 \\ \sum_{k=-3}^{k=3} y_k k = 28a_1 + 169a_3 \\ \sum_{k=-3}^{k=3} y_k k^2 = 28a_0 + 169a_2 \\ \sum_{k=-3}^{k=3} y_k k^3 = 169a_1 + 1588a_3 \end{cases}$$

Определим коэффициенты разложения

$$\begin{cases} \sum_{k=-3}^{k=3} y_k = 9a_0 + 28a_2 \\ \sum_{k=-3}^{k=3} y_k k = 28a_1 + 169a_3 \\ \sum_{k=-3}^{k=3} y_k k^2 = 28a_0 + 169a_2 \\ \sum_{k=-3}^{k=3} y_k k^3 = 169a_1 + 1588a_3 \end{cases} \rightarrow A = \begin{pmatrix} 9 & 0 & 28 & 0 \\ 0 & 28 & 0 & 169 \\ 28 & 0 & 169 & 0 \\ 0 & 169 & 0 & 1588 \end{pmatrix}, \quad B = \begin{pmatrix} \sum_{k=-3}^{k=3} y_k \\ \sum_{k=-3}^{k=3} y_k k \\ \sum_{k=-3}^{k=3} y_k k^2 \\ \sum_{k=-3}^{k=3} y_k k^3 \end{pmatrix}$$

$$\begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{pmatrix} = A^{-1} B = A^{-1} \begin{pmatrix} \sum_{k=-3}^{k=3} y_k \\ \sum_{k=-3}^{k=3} y_k k \\ \sum_{k=-3}^{k=3} y_k k^2 \\ \sum_{k=-3}^{k=3} y_k k^3 \end{pmatrix}$$

Для решения этого уравнения будем использовать *Mathcad*

$$\begin{aligned}
A &:= \begin{pmatrix} 7 & 0 & \sum_{k=-3}^3 k^2 & 0 \\ 0 & \sum_{k=-3}^3 k^2 & 0 & \sum_{k=-3}^3 k^4 \\ \sum_{k=-3}^3 k^2 & 0 & \sum_{k=-3}^3 k^4 & 0 \\ 0 & \sum_{k=-3}^3 k^4 & 0 & \sum_{k=-3}^3 k^6 \end{pmatrix} \\
A &\rightarrow \begin{pmatrix} 7 & 0 & 28 & 0 \\ 0 & 28 & 0 & 196 \\ 28 & 0 & 196 & 0 \\ 0 & 196 & 0 & 1588 \end{pmatrix} \quad B := \begin{pmatrix} \sum_{k=-3}^3 y_k \\ \sum_{k=-3}^3 (y_k \cdot k) \\ \sum_{k=-3}^3 (y_k \cdot k^2) \\ \sum_{k=-3}^3 (y_k \cdot k^3) \end{pmatrix} \\
\begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{pmatrix} &:= A^{-1} \cdot B \rightarrow \begin{pmatrix} \frac{y_0}{3} + \frac{2 \cdot y_{-1}}{7} + \frac{2 \cdot y_1}{7} + \frac{y_{-2}}{7} + \frac{y_2}{7} - \frac{2 \cdot y_{-3}}{21} - \frac{2 \cdot y_3}{21} \\ \frac{29 \cdot y_1}{126} - \frac{29 \cdot y_{-1}}{126} - \frac{67 \cdot y_{-2}}{252} + \frac{67 \cdot y_2}{252} + \frac{11 \cdot y_{-3}}{126} - \frac{11 \cdot y_3}{126} \\ \frac{5 \cdot y_{-3}}{84} - \frac{y_{-1}}{28} - \frac{y_1}{28} - \frac{y_0}{21} + \frac{5 \cdot y_3}{84} \\ \frac{y_{-1}}{36} - \frac{y_1}{36} + \frac{y_{-2}}{36} - \frac{y_2}{36} - \frac{y_{-3}}{36} + \frac{y_3}{36} \end{pmatrix}
\end{aligned}$$

Запишем окончательную формулу для  $a_0$

$$\begin{aligned}
a_0 &= \frac{1}{3}y_0 + \frac{2}{7}(y_{-1} + y_1) + \frac{1}{7}(y_{-2} + y_2) - \frac{2}{21}(y_{-3} + y_3) = \\
&\frac{7}{21}y_0 + \frac{6}{21}(y_{-1} + y_1) + \frac{3}{21}(y_{-2} + y_2) - \frac{2}{21}(y_{-3} + y_3)
\end{aligned}$$

Выпишем весовые коэффициенты для полинома третьей степени в таблицу

$$-\frac{2}{21}; \frac{3}{21}; \frac{6}{21}; \frac{7}{21}; \frac{6}{21}; \frac{3}{21}; -\frac{2}{21}$$

Учитывая симметрию их можно записать в виде

$$\frac{1}{21}(-2; 3; 6; 7);$$

Домашняя работа получить коэффициенты для  $n=5$ ,  $N=9$

1. Статистическая обработка
2. МНК
3. Регрессия
4. Нелинейная регрессия
5. Аппроксимация Паде (экстраполяция, прогнозирование)
6. Прогнозирование.