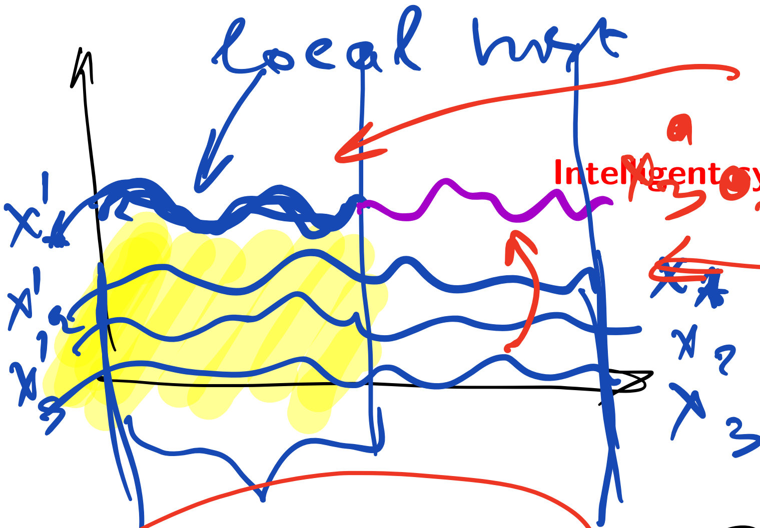
# Mathematical methods of forecasting

Intelligent systems, Phystech



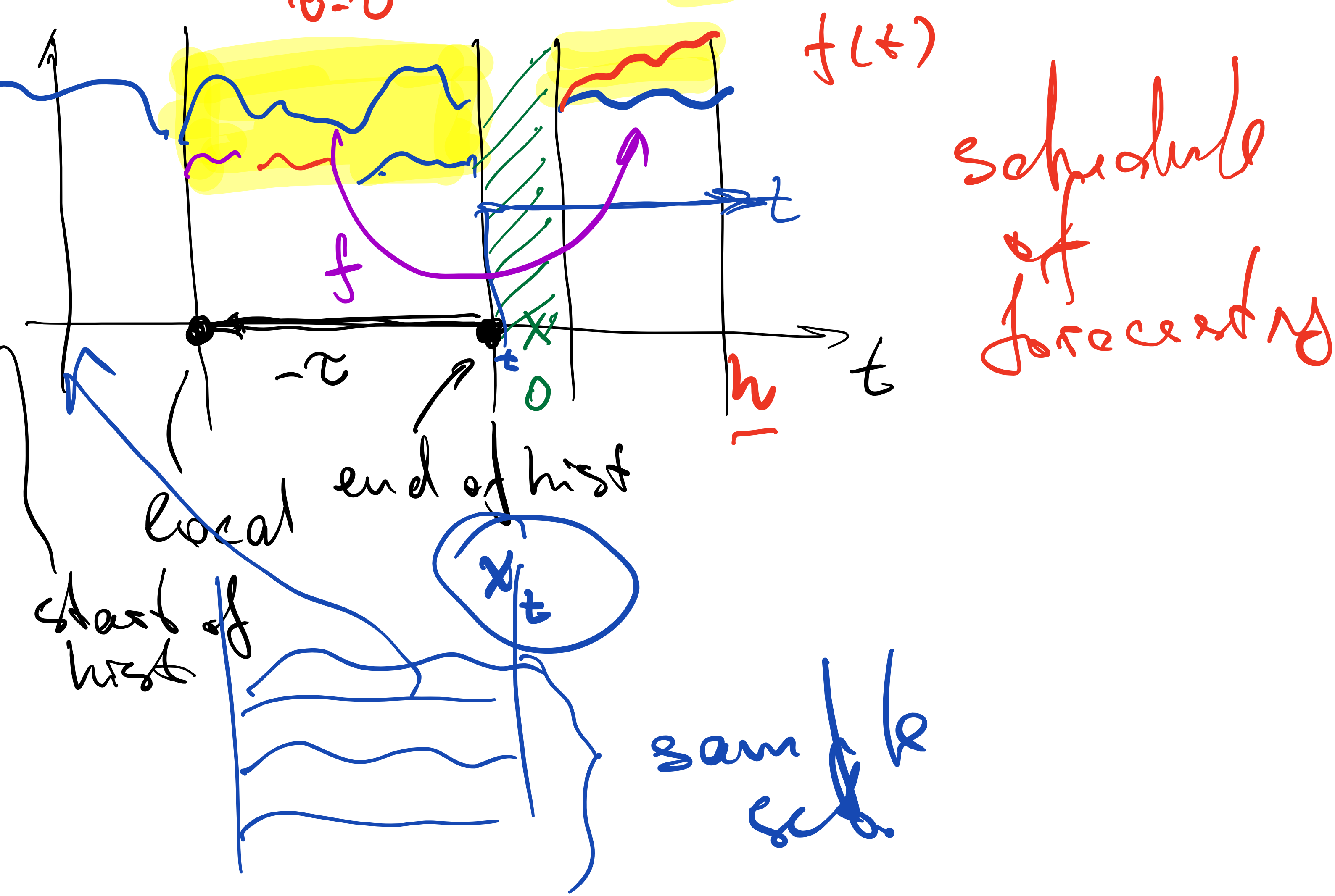

$$\hat{x} = \frac{x_1 + x_2 + x_3}{3}$$

$$s_t = \alpha x_t + (1-\alpha) s_{t-1}$$

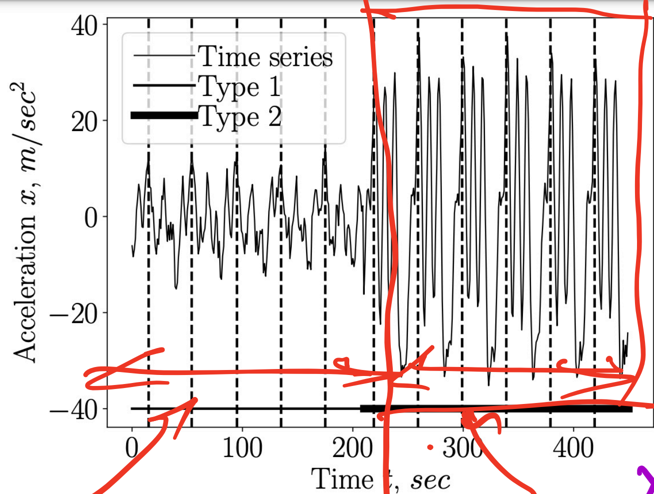
$$s_t = \alpha \sum (1-\alpha)^{\tau} x_{t-\tau}$$

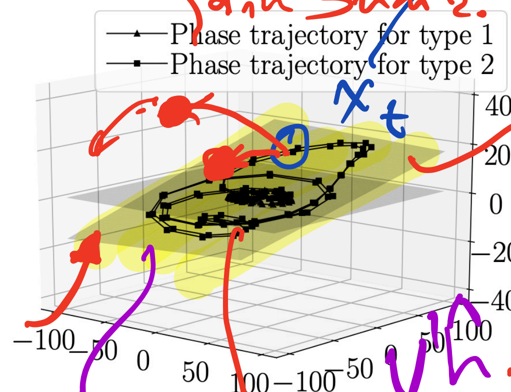
$$\hat{x} = \exp\left(\delta x^{T} \Sigma \delta x\right)$$

**2022**


f(t)
schedule of forecasting
f
-τ
t
0
h
t
end of hist
local
start of hist
x_t
sample set

# SSA




dim 1000

dim 3 или 2.

$x_t$

$Vx$

шаг

анн

бер.

бег.

нар

кон

$q$

$x' = V'h \Rightarrow q = Vh$

$\Sigma = \frac{1}{n} X^T X$

$$\underset{m\times n}{X} = \underset{m\times n}{U}\ \underset{n\times n}{\Lambda}\ \underset{n\times n}{V^T}$$

$$\hat{X} = \sum_{k=1} X_k$$

$$X_k = \lambda_k u_k v_k^T$$

$V \Lambda^2 V^T V \Lambda V^T$

$\Sigma V = \Lambda^2 V$

$a = [1, 3, 5]$

$[1\ 0\ 1\ 0\ 1]$

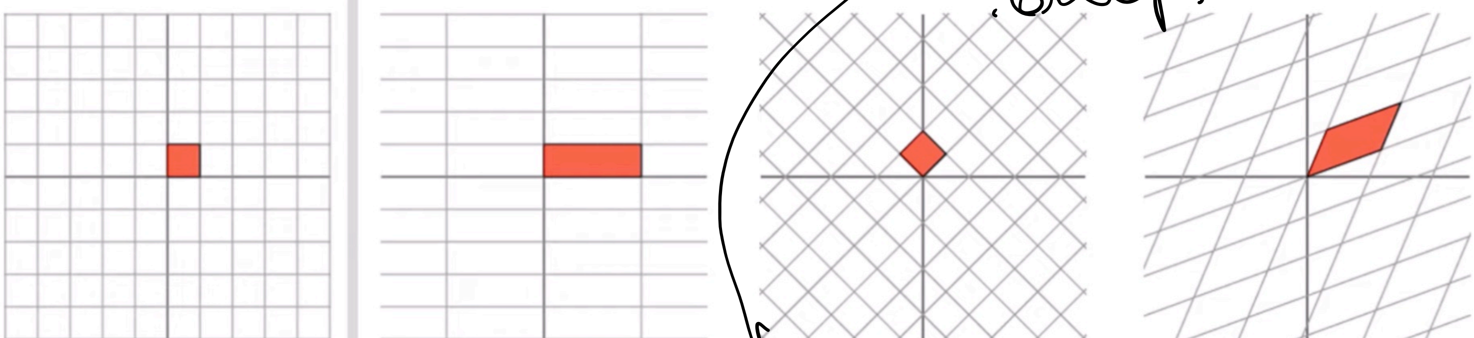
$\hat{x} = V^T h$ — как при SSA.

Linear Maps (geometrically) are spatial transforms that...

1. Keep gridlines parallel
2. Keep gridlines evenly spaced
3. Keep the origin stationary

$X^T X \propto \Sigma$

$u v^T = u \otimes v$



(Images taken from the American Math Society.)

SIAM.

$X = U \Lambda V^T$

rank$(X) = r$

$\lambda_1 \geq \dots \lambda_n \geq 0$

Рассмотрим квадратичный алгоритм решения этой задачи. Найдем последовательно векторы $\mathbf{u}_k, \mathbf{v}_k$ и сингулярные числа $\lambda_k$ для $k = 1, \ldots, r$. В качестве этих векторов берутся нормированные значения векторов $\mathbf{a}_k$ и $\mathbf{b}_k$, соответственно

$X = X_1 + \ldots$

$$\underset{m\times 1}{\mathbf{u}_k} = \frac{\mathbf{a}_k}{\|\mathbf{a}_k\|} \quad \text{и} \quad \underset{1\times n}{\mathbf{v}_k} = \frac{\mathbf{b}_k}{\|\mathbf{b}k\|}.$$

$X^T a, Xb$

Векторы $\mathbf{a}_k$ и $\mathbf{b}_k$ находятся как пределы последовательностей векторов $\{\mathbf{a}_{k_s}\}$ и $\{\mathbf{b}_{k_s}\}$, соответственно

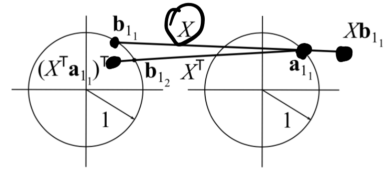
$$\mathbf{a}_k = \lim_{s\to\infty}(\mathbf{a}_{k_s}) \ \text{ и } \mathbf{b}_k = \lim_{s\to\infty}(\mathbf{b}_{k_s}).$$

$X, X^T$

Сингулярное число $\lambda_k$ находится как произведение норм векторов

$$\lambda_k = \|\mathbf{a}_k\| \cdot \|\mathbf{b}_k\|.$$

A




B.

Рис. 13. Итеративная процедура оценивания сингулярных векторов.

Процедура нахождения последовательностей векторов $\mathbf{a}_{k_s}, \mathbf{b}_{k_s}$ $\mathbf{u}_k, \mathbf{v}_k$ начинается с выбора наибольшей по норме строки $\mathbf{b}_{1_1}$ матрицы $\mathbf{X}$. Для $k = 1$ формулы нахождения векторов $\mathbf{a}_{1_s}, \mathbf{b}_{1_s}$ имеют вид:

$$\underset{m\times 1}{\mathbf{a}_{1_s}} = \frac{\mathbf{X}\mathbf{b}_{1_s}^\mathsf{T}}{\mathbf{b}_{1_s}\mathbf{b}_{1_s}^\mathsf{T}}, \quad \underset{1\times n}{\mathbf{b}_{1_{s+1}}} = \frac{\mathbf{a}_{1_s}^\mathsf{T}\mathbf{X}}{\mathbf{a}_{1_s}^\mathsf{T}\mathbf{a}_{1_s}}, \quad s = 1, 2, \cdots.$$

Для вычисления векторов $\mathbf{u}_k, \mathbf{v}_k$ при $k = 2, \ldots, r$ используется вышеприведенная формула, с той разницей, что матрица $\mathbf{X}$ заменяется на скорректированную на $k$-м шаге матрицу $\mathbf{X}_{k+1} = \mathbf{X}_k - \mathbf{u}_k\lambda_k\mathbf{v}_k$. На рисунке **??** показаны две итерации, $s = 1, 2$, первого шага $k = 1$ упрощенной процедуры нахождения сингулярного разложения.

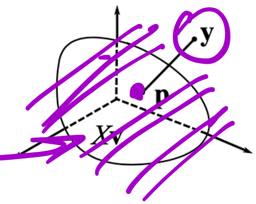
Требуется минимизировать евклидово расстояние от вектора $\mathbf{y}$ до вектора $\mathbf{Xw}$. Этот вектор лежит в пространстве столбцов матрицы $\mathbf{X}$, так как $\mathbf{Xw}$ — это линейная комбинация столбцов этой матрицы с коэффициентами $w_1, \ldots, w_n$. Задача оценки $\mathbf{w}$ эквивалентна задаче нахождения точки $\mathbf{p} = \mathbf{Xw}$, ближайшей к $\mathbf{y}$ и находящейся в пространстве столбцов матрицы $\mathbf{X}$. Следовательно, вектор $\mathbf{p}$ должен быть проекцией $\mathbf{y}$ на пространство столбцов, вектор регрессионных остатков $\mathbf{Xw} - \mathbf{y}$ должен быть ортогонален этому пространству. Рассмотрим произвольный вектор $\mathbf{Xv}$, ортогональный вектору регрессионных остатков $\mathbf{Xw} - \mathbf{y}$:

$$(\mathbf{Xv})^\mathsf{T}(\mathbf{Xw} - \mathbf{y}) = \mathbf{v}^\mathsf{T}(\mathbf{X}^\mathsf{T}\mathbf{Xw} - \mathbf{X}^\mathsf{T}\mathbf{y}) = 0.$$

Так как это равенство должно быть справедливо для произвольного вектора $\mathbf{v}$, то $\mathbf{X}^\mathsf{T}\mathbf{Xw} - \mathbf{X}^\mathsf{T}\mathbf{y} = 0$, см. рис. ??. Если столбцы матрицы $\mathbf{X}$ линейно независимы, то матрица $\mathbf{X}^\mathsf{T}\mathbf{X}$ обратима и уравнение имеет единственное решение относительно параметров

$$\mathbf{w} = (\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\mathbf{X}^\mathsf{T}\mathbf{y}. \tag{38}$$

Xv =




$\hat{x} = w^T x$

Рис. 6. Проекция вектора зависимой переменной на пространство столбцов матрицы плана.

Проекция вектора $\mathbf{y}$ на пространство столбцов матрицы $\mathbf{X}$ имеет вид

$$\mathbf{p} = \mathbf{Xw} = \mathbf{X}(\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\mathbf{X}^\mathsf{T}\mathbf{y} = \mathbf{Py}.$$

Матрица $\mathbf{P} = \mathbf{X}(\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\mathbf{X}^\mathsf{T}$ называется матрицей проектирования. Она она идемпотентна, $\mathbf{P}^2 = \mathbf{P}$, и симметрична, $\mathbf{P}^\mathsf{T} = \mathbf{P}$.

SSA can be used as a model-free technique so that it can be applied to arbitrary time series including non-stationary time series. The basic aim of SSA is to decompose the time series into the sum of interpretable components such as trend, periodic components and noise with no a-priori assumptions about the parametric form of these components.

Consider a real-valued time series $\mathbb{X} = (x_1, \ldots, x_N)$ of length $N$. Let $L \ \ (1 < L < N)$ be some integer called the *window length* and $K = N - L + 1$.

### Main algorithm

**1st step: Embedding.**

Form the *trajectory matrix* of the series $\mathbb{X}$, which is the $L \times K$ matrix

$$\mathbf{X} = [X_1 : \ldots : X_K] = (x_{ij})_{i,j=1}^{L,K} = \begin{bmatrix} x_1 & x_2 & x_3 & \ldots & x_K \\ x_2 & x_3 & x_4 & \ldots & x_{K+1} \\ x_3 & x_4 & x_5 & \ldots & x_{K+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & x_{L+2} & \ldots & x_N \end{bmatrix}$$

where $X_i = (x_i, \ldots, x_{i+L-1})^{\mathrm{T}} \quad (1 \le i \le K)$ are *lagged vectors* of size $L$. The matrix $\mathbf{X}$ is a Hankel matrix which means that $\mathbf{X}$ has equal elements $x_{ij}$ on the anti-diagonals $i + j =$ const.

**2nd step: Singular Value Decomposition (SVD).**

Perform the singular value decomposition (SVD) of the trajectory matrix $\mathbf{X}$. Set $\mathbf{S} = \mathbf{X}\mathbf{X}^{\mathrm{T}}$ and denote by $\lambda_1, \ldots, \lambda_L$ the *eigenvalues* of $\mathbf{S}$ taken in the decreasing order of magnitude ($\lambda_1 \geq \ldots \geq \lambda_L \geq 0$) and by $U_1, \ldots, U_L$ the orthonormal system of the *eigenvectors* of the matrix $\mathbf{S}$ corresponding to these eigenvalues.

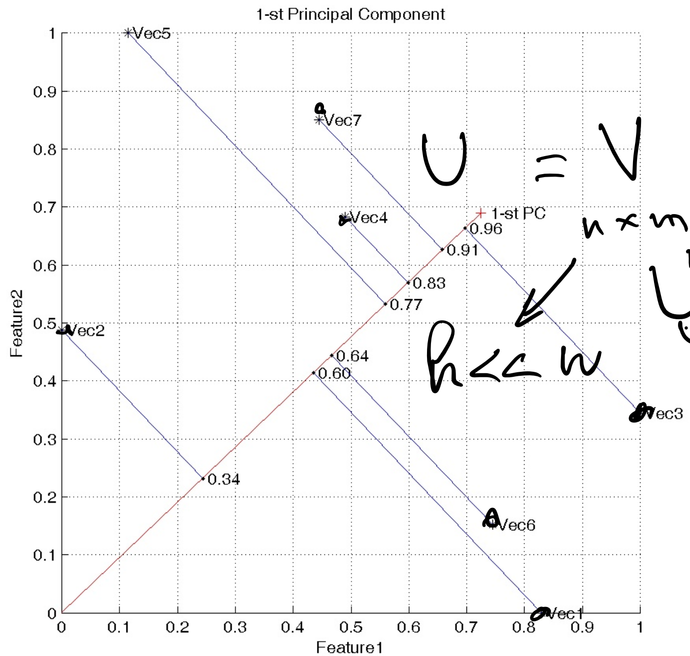
Set $d = \mathop{\mathrm{rank}} \mathbf{X} = \max\{i, \ \mbox{such that} \ \lambda_i > 0\}$ (note that $d = L$ for a typical real-life series) and $V_i = \mathbf{X}^{\mathrm{T}} U_i / \sqrt{\lambda_i}$ $(i = 1, \ldots, d)$. In this notation, the SVD of the trajectory matrix $\mathbf{X}$ can be written as

$$\mathbf{X} = \mathbf{X}_1 + \ldots + \mathbf{X}_d,$$

where

$$\mathbf{X}_i = \sqrt{\lambda_i} U_i V_i^{\mathrm{T}}$$

are matrices having rank 1; these are called *elementary matrices*. The collection $(\sqrt{\lambda_i}, U_i, V_i)$ will be called the $i$th *eigentriple* (abbreviated as ET) of the SVD. Vectors $U_i$ are the left singular vectors of the matrix $\mathbf{X}$, numbers $\sqrt{\lambda_i}$ are the singular values and provide the singular spectrum of $\mathbf{X}$; this gives the name to SSA. Vectors $\sqrt{\lambda_i} V_i = \mathbf{X}^{\mathrm{T}} U_i$ are called vectors of principal components (PCs).




$$\underset{n \times m}{U} = V \underset{m \times n}{X^T}$$

$n << m$

*Проекция $p$ точки $b$ на прямую, определенную вектором $a$, задается формулой*

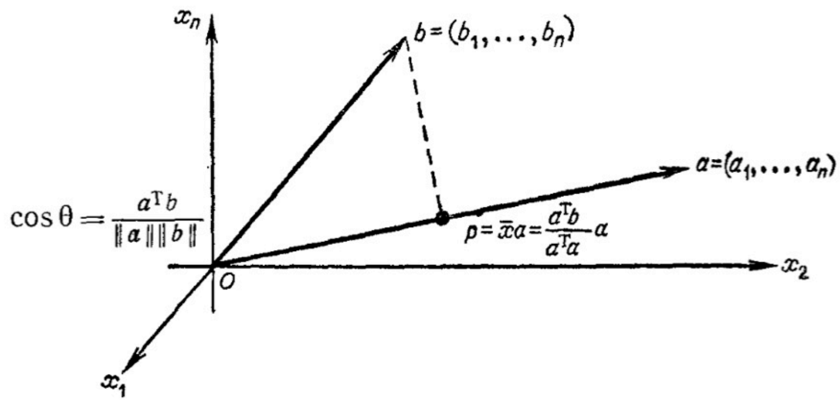
$$p = \frac{a^{\mathrm{T}} b}{a^{\mathrm{T}} a} a.$$

*Расстояние (в квадрате) от этой точки до прямой равняется*

$$\left\| b - \frac{a^{\mathrm{T}} b}{a^{\mathrm{T}} a} a \right\|^2 = b^{\mathrm{T}} b - 2 \frac{(a^{\mathrm{T}} b)^2}{a^{\mathrm{T}} a} + \left( \frac{a^{\mathrm{T}} b}{a^{\mathrm{T}} a} \right)^2 a^{\mathrm{T}} a =$$

$$= \frac{(b^{\mathrm{T}} b)(a^{\mathrm{T}} a) - (a^{\mathrm{T}} b)^2}{(a^{\mathrm{T}} a)}.$$

Это позволяет нам повторить рис. 3.1 уже с указанием формулы для определения точки $p$ (рис. 3.3).




Проекция вектора $b$ на вектор $a$.

$$
\begin{tikzcd}
(-\lambda^2\nabla_{\mathbf{w}}S+\mathbf{w}^*) \arrow[r, "T"] & \hat{\mathbf{w}},\mathbf{x} \arrow[r, "f"] & \hat{y} \arrow[r] & S(\mathbf{w}|\mathrm{y},\hat{\mathrm{y}}) \arrow[lll, bend right] \\
t \arrow[r] & i \arrow[r] & (\mathbf{x},y) \arrow[ul] \arrow[ur] &
\end{tikzcd}
$$

$f$ is the forecasting model,

$S$ is the criterion,

$T$ is an optimization algorithm,

$\hat{\mathbf{w}}$ is some solution,

$$\hat{\mathbf{w}} = \arg\min S(\mathbf{w}|y,f).$$

[1] These notations are equivalent: $x_i, x(i), i \to x$.

[1] *