Нейронные сети с радиальными базисными функциями (RBF)

RBF – ***функции, радиально изменяющиеся вокруг некоторого центра, заданного вектором* С, *и принимающие ненулевые значения в окрестности этого центра.***

Аргумент RBF - расстояние между входным вектором X и центром C:

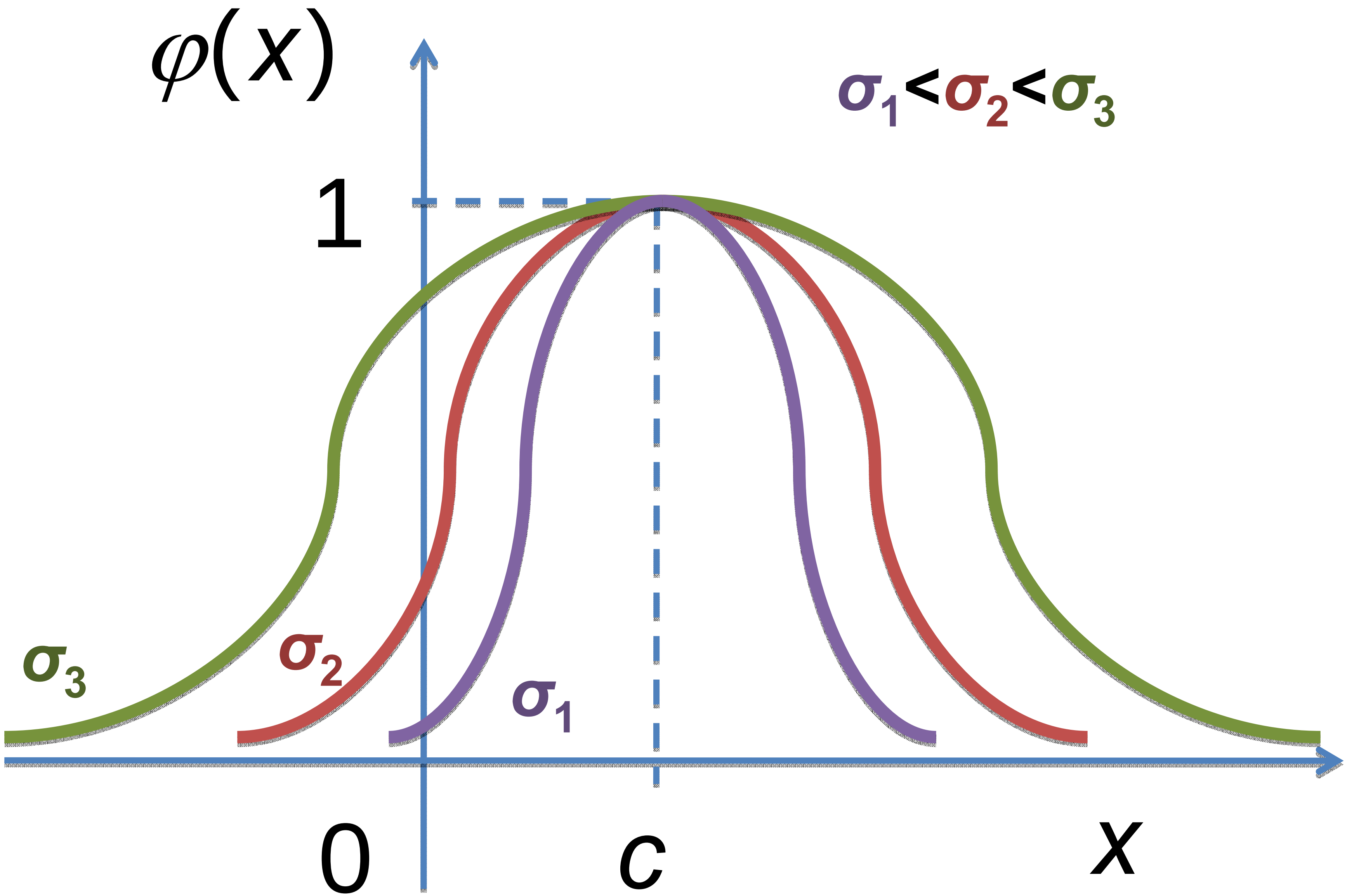
$$\varphi = \varphi\left(\|\mathbf{X} - \mathbf{C}\|\right)$$

где

$$\|\mathbf{X} - \mathbf{C}\| = \sqrt{\left(x_1 - c_1\right)^2 + \ldots + \left(x_N - c_N\right)^2}$$

Гауссовы RBF

$$\varphi(\mathbf{X}) = \exp\left(-\frac{\|\mathbf{X}-\mathbf{C}\|^2}{2\sigma^2}\right)$$



$\varphi(x)$
$\sigma_1 < \sigma_2 < \sigma_3$
1
σ_3
σ_2
σ_1
0
c
x

Другие типы RBF

$$\text{функция Коши}: \quad \left(1+\|\mathbf{X}-\mathbf{C}\|^2\right)^{-1}$$

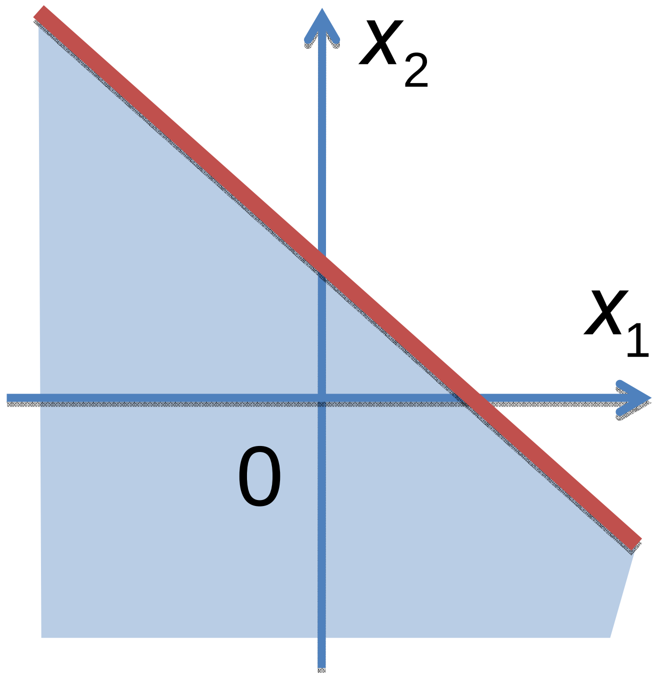
$$thin-plate\ splines: \quad \|\mathbf{X}-\mathbf{C}\|\ln\left(\|\mathbf{X}-\mathbf{C}\|\right)$$

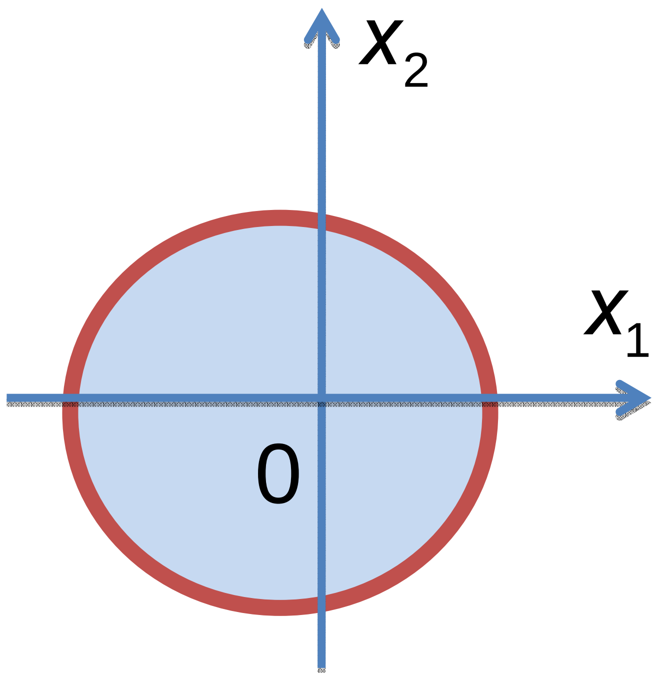
$$\text{линейный сплайн}: \quad \|\mathbf{X}-\mathbf{C}\|$$

$$\text{кубический сплайн}: \quad \|\mathbf{X}-\mathbf{C}\|^3$$

$$\text{мультиквадрика}: \quad \sqrt{1+\frac{\|\mathbf{X}-\mathbf{C}\|^2}{\sigma^2}}$$

Разбиение пространства на классы





$$w_1 x_1 + w_2 x_2 + w_0 = 0$$

$$(x_1 - c_1)^2 + (x_2 - c_2)^2 - R^2 = 0$$

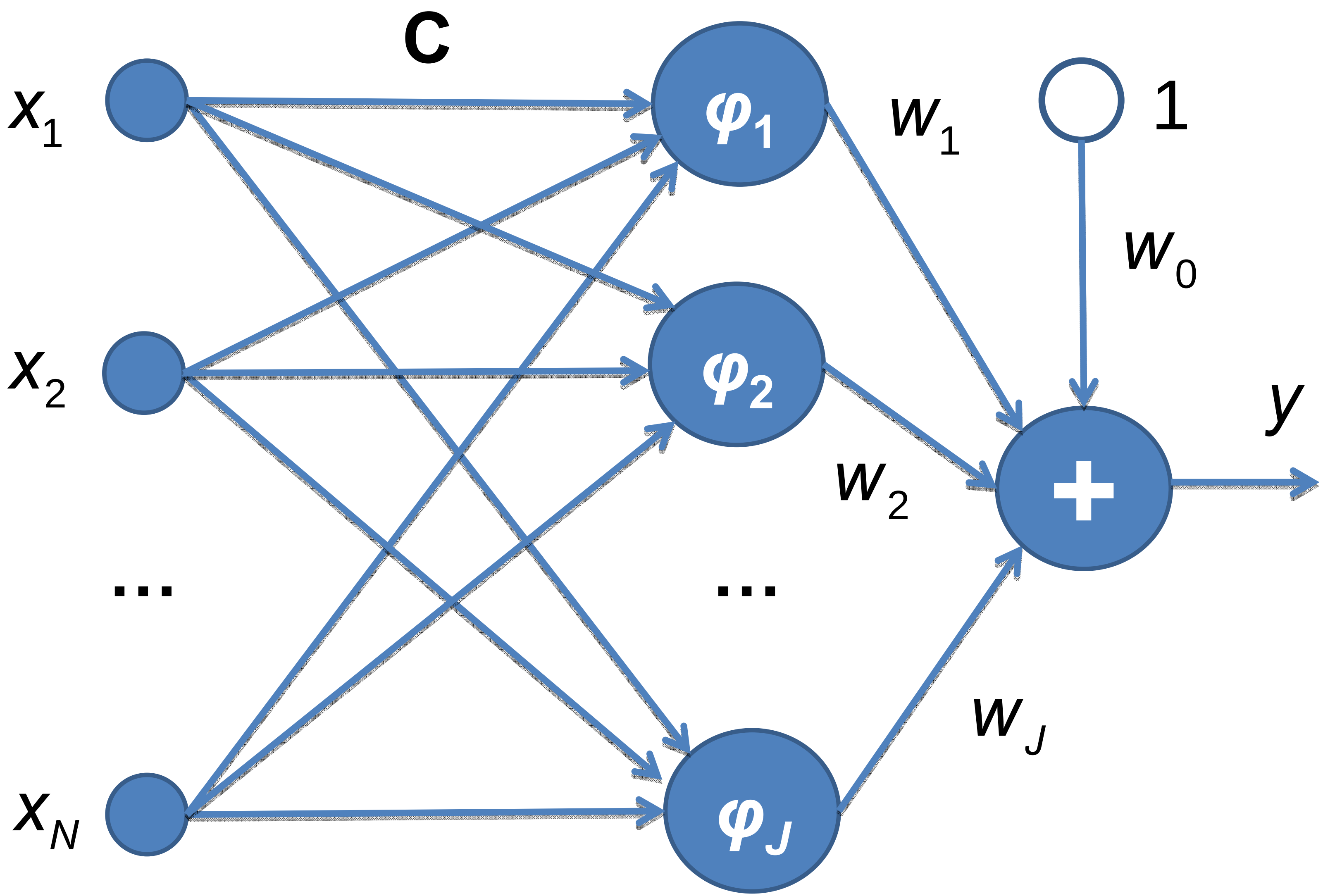
Сигмоидальный нейрон

Нейрон RBF

Основное преимущество RBF

– легкость разбиения пространства на классы → снижение количества скрытых слоев.

N вх. → J RBF → 1 вых.


x_1
x_2
...
x_N
C
φ_1
φ_2
...
φ_J
w_1
w_2
w_J
1
w_0
+
y

Алгоритм функционирования RBF

$$\mathrm{net}_j = \left\| \mathbf{X} - \mathbf{C}_j \right\| =$$

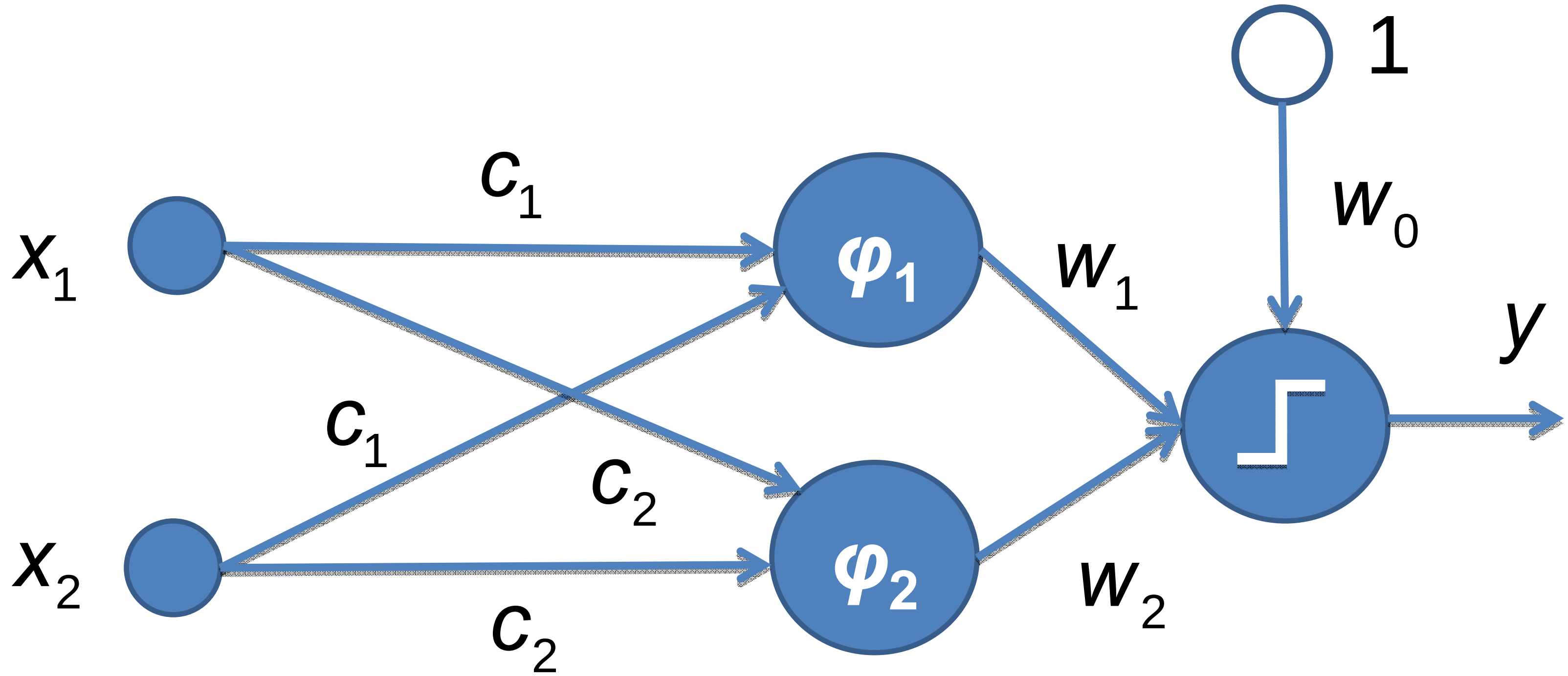
$$= \sqrt{\left(x_1 - c_{j1}\right)^2 + \ldots + \left(x_N - c_{jN}\right)^2}$$

$$\varphi_j = \varphi\left(-\mathrm{net}_j^2\right)$$

$$y = \sum_{j=1}^{J} w_j \varphi_j + w_0$$

Пример (XOR)

y	0	1	1	0
x_1	0	0	1	1
x_2	0	1	0	1



$$\varphi_1 = \exp\left(-\mathrm{net}_1^2\right) =$$

$$= \sqrt{\left(x_1 - c_1\right)^2 + \left(x_2 - c_1\right)^2}$$

$$\varphi_2 = \exp\left(-\mathrm{net}_2^2\right) =$$

$$= \sqrt{\left(x_1 - c_2\right)^2 + \left(x_2 - c_2\right)^2}$$

Пусть

$$c_1 = 1, \quad c_2 = 0$$

Тогда

x_1	0	0	1	1
x_2	0	1	0	1
φ_1	e^{-2}	e^{-1}	e^{-1}	1
φ_2	1	e^{-1}	e^{-1}	e^{-2}

Пусть

$$w_1 = w_2 = 1, \quad w_0 = -1$$

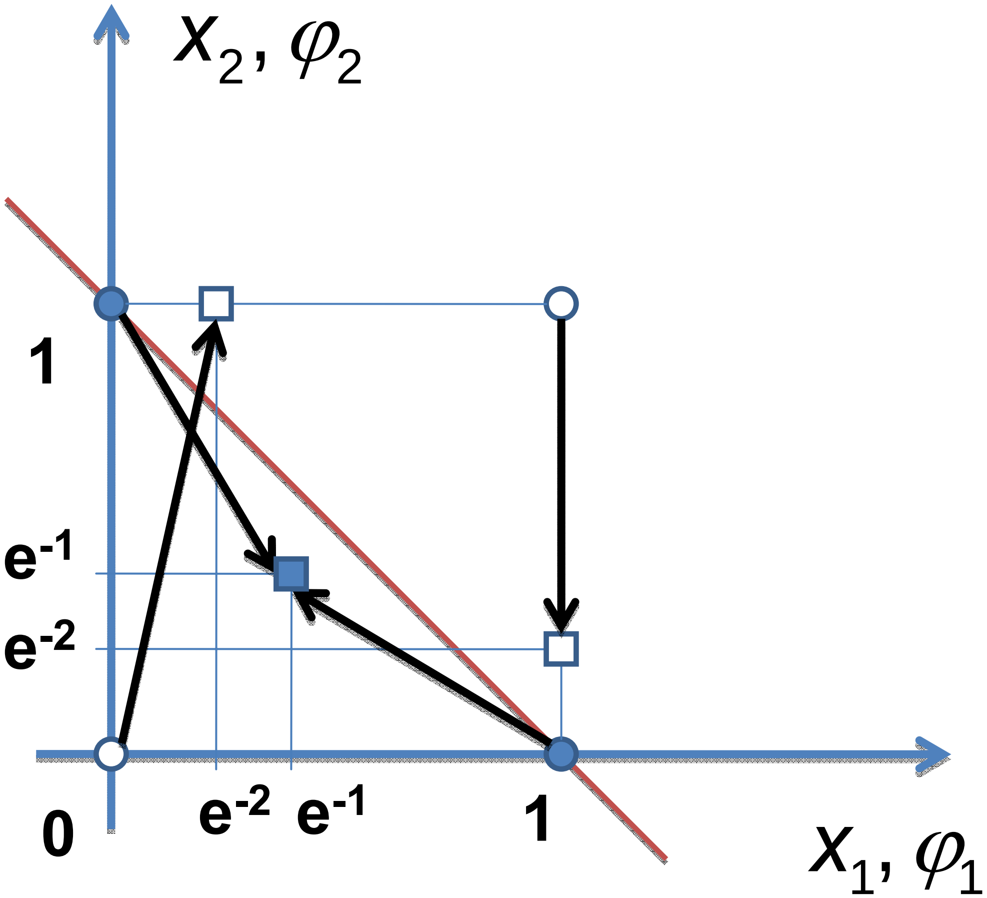
Тогда дискриминантная прямая

$$w_1\varphi_1 + w_2\varphi_2 + w_0 = 0$$

имеет вид

$$\varphi_1 + \varphi_2 - 1 = 0 \quad \textit{или} \quad \varphi_2 = 1 - \varphi_1$$

Линейная разделимость по φ_1, φ_2:



$e^{-1} \sim 0.368$
$e^{-2} \sim 0.135$

Алгоритмы обучения RBF

Обучающая выборка:

$$\mathbf{X}^{(q)}, \quad q = \overline{1,Q}$$

Целевой выход:

$$t_q, \quad q = \overline{1,Q}$$

Система:

$$\begin{bmatrix} 1 & \left\|\mathbf{X}^{(1)}-\mathbf{C}_1\right\| & \dots & \left\|\mathbf{X}^{(1)}-\mathbf{C}_J\right\| \\ 1 & \left\|\mathbf{X}^{(2)}-\mathbf{C}_1\right\| & \dots & \left\|\mathbf{X}^{(2)}-\mathbf{C}_J\right\| \\ 1 & \dots & \dots & \dots \\ 1 & \left\|\mathbf{X}^{(Q)}-\mathbf{C}_1\right\| & \dots & \left\|\mathbf{X}^{(Q)}-\mathbf{C}_J\right\| \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_J \end{bmatrix} = \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_Q \end{bmatrix}$$

или

$$\mathbf{\Phi w} = \mathbf{t}$$

Если

$$J = Q$$

то

$$\mathbf{w} = \mathbf{\Phi}^{-1}\mathbf{t}$$

Плохо, т.к. при $J \gg 1$ усложняется структура НС, возникает переобучение НС

Поэтому обычно

$$J << Q$$

→ матрица **Φ** – прямоугольная. Тогда

$$\mathbf{w} = \mathbf{\Phi}^{+}\mathbf{t}$$

где **Φ⁺** – псевдообратная матрица

$$\mathbf{\Phi}^{+} = \left(\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}\right)^{-1}\mathbf{\Phi}^{\mathrm{T}}$$

Веса w_j определяются из условия СКО → min:

$$\varepsilon = \sum_{q=1}^{Q}\left[\sum_{j=1}^{J} w_j \varphi\left(\left\| \mathbf{X}^{(q)} - \mathbf{C}_j \right\|\right) + w_0 - t^{(q)}\right]$$

или методом Видроу-Хоффа.

$$\varphi_j(\mathbf{X}) = \exp\left(-\frac{\left\|\mathbf{X} - \mathbf{C}_j\right\|^2}{2\sigma_j^2}\right)$$

Выбор параметров

$$\mathbf{C}_j, \sigma_j \quad - ?$$

1) При *J*=*Q*

$$\mathbf{c}_j = \mathbf{X}^{(j)} \quad (j = \overline{1,Q})$$

$\boldsymbol{\sigma}_j$ – часть пространства, в которой $\mathbf{X}^{(q)}$ была охвачена ФА, например:

$$\sigma_j = \sqrt{\frac{1}{P}\sum_{i=1}^{P}\left\|\mathbf{c}_j - \mathbf{c}_i\right\|}, \quad P = 3 \ldots 5$$

Затем $\quad \mathbf{w} = \boldsymbol{\Phi}^{-1}\mathbf{t}$

2) При $J \ll Q$

- предварительная кластеризация векторов $\mathbf{X}^{(q)}$
- определение центров кластеров $\mathbf{C}_j$
- определение $\boldsymbol{\sigma}_j$
- определение весов $\mathbf{w}$

(min LSE, Видроу-Хофф, $\mathbf{w}=\mathbf{\Phi}^{+}\mathbf{t}$, …)

Преимущества НС RBF

- простота топологии
- простота обучения
- частично решается проблема локальных min

Сложности

- J - ?
- C_j - ?
- σ_j - ?

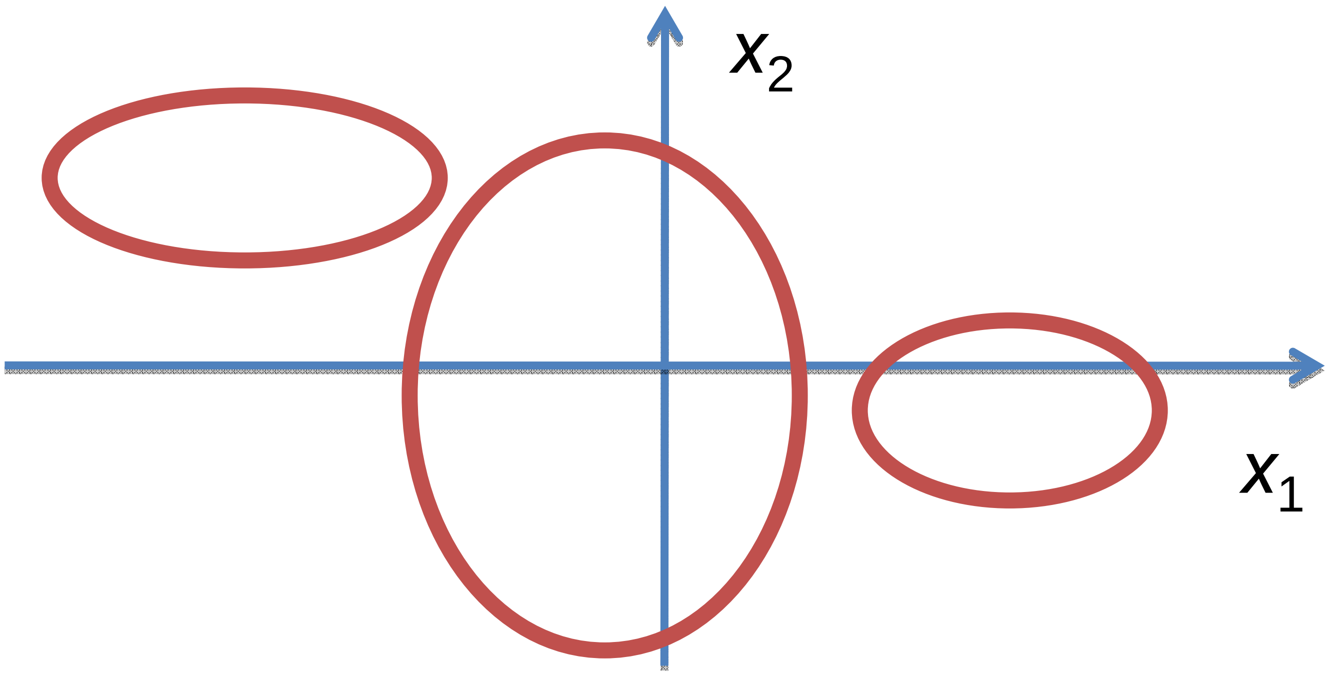
Недостаток:

RBF-сети неприменимы к задачам экстраполяции

<u>Асимметричные RBF</u>

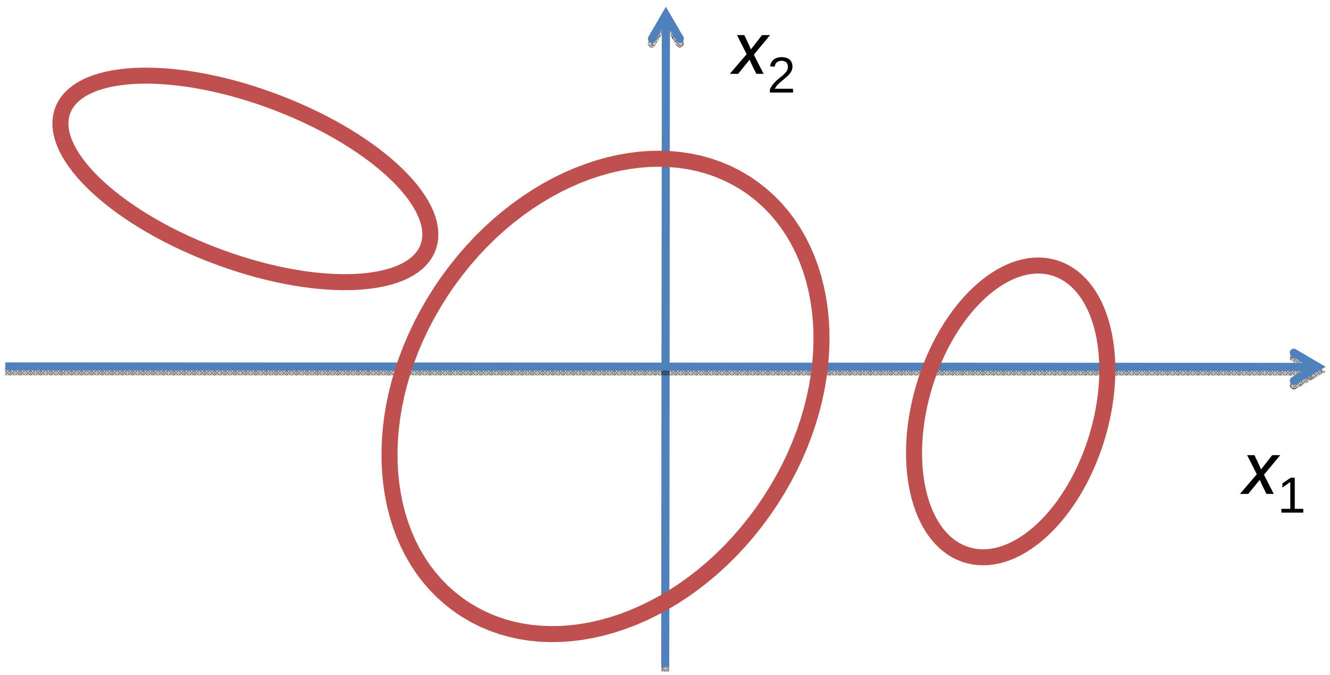
(не все направления от центра RBF равноправны - анизотропия координат)

$$\varphi_j(x_1,\ldots,x_k) = \exp\left(-\sum_{i=1}^{k} \alpha_{ij}(x_i - c_{ij})^2\right)$$

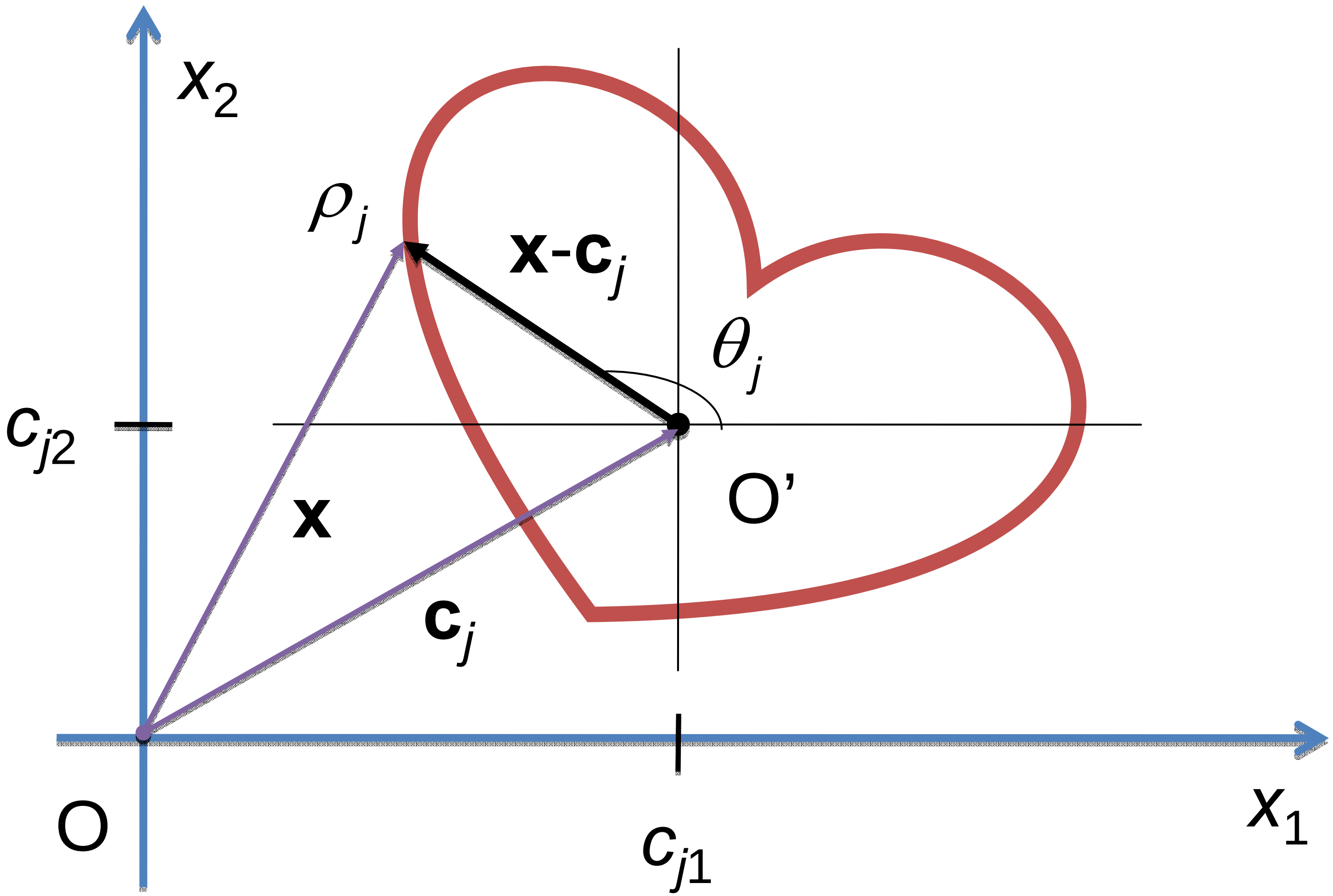


$$\varphi_j(x_1,\ldots,x_k) =$$

$$= \exp\left(-\sum_{i=1}^{k}\sum_{l=1}^{k}\alpha_{ijl}(x_i - c_{ij})(x_l - c_{lj})\right)$$



Произвольная асимметрия вдоль лучей из c_j (звездность)



Локальные полярные координаты:

$$\begin{cases} \rho_j = \left\| \mathbf{x} - \mathbf{c}_j \right\| \\ \theta_j = \arg(\mathbf{x} - \mathbf{c}_j) \end{cases}$$

Асимметричная RBF:

$$\varphi_j = \varphi(\rho - \mathbf{c}_j, \theta_j)$$

Пусть $$\rho = f(\theta)$$

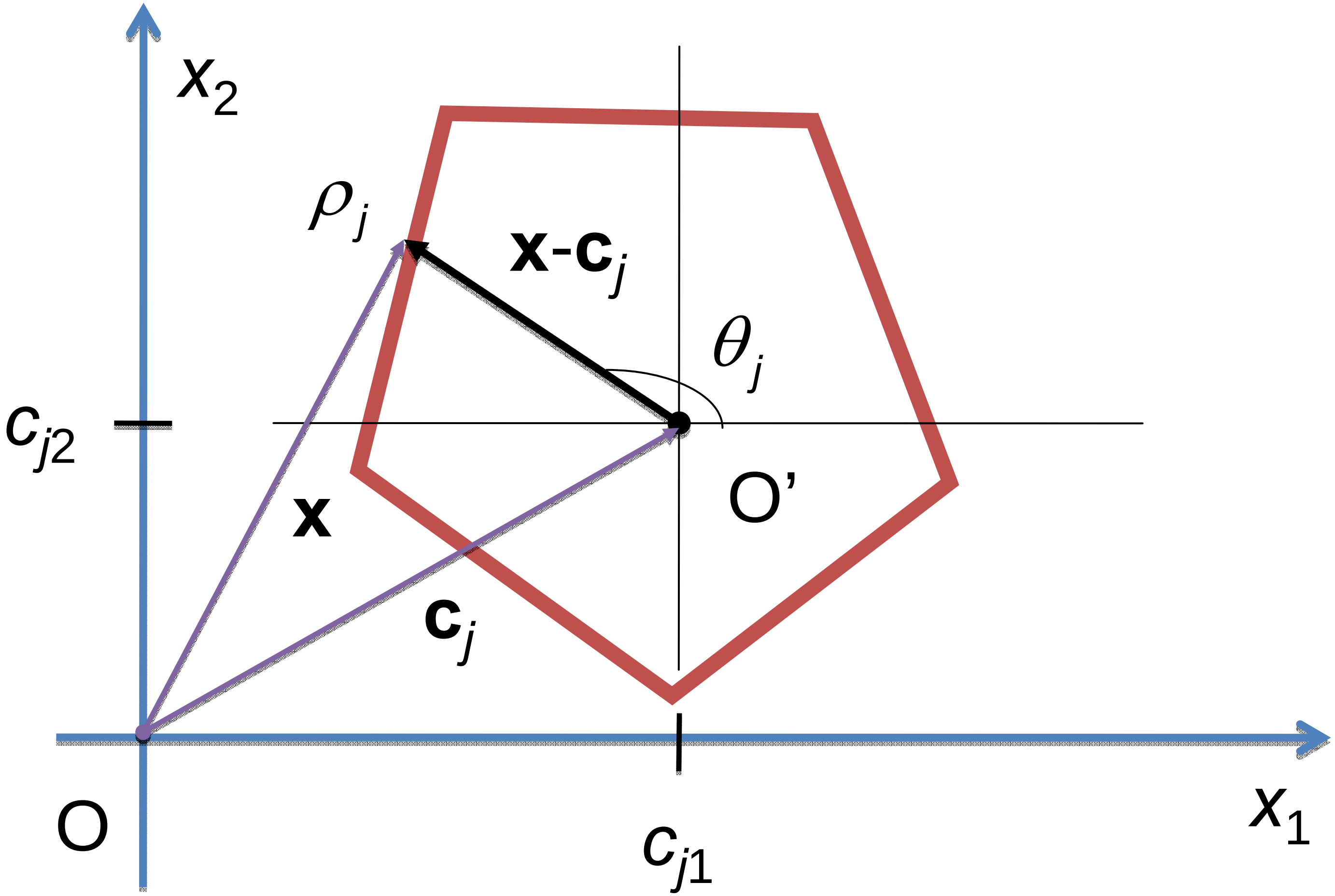
тогда

$$\varphi = \left(1 - \frac{\rho}{f(\theta)}\right)_+$$

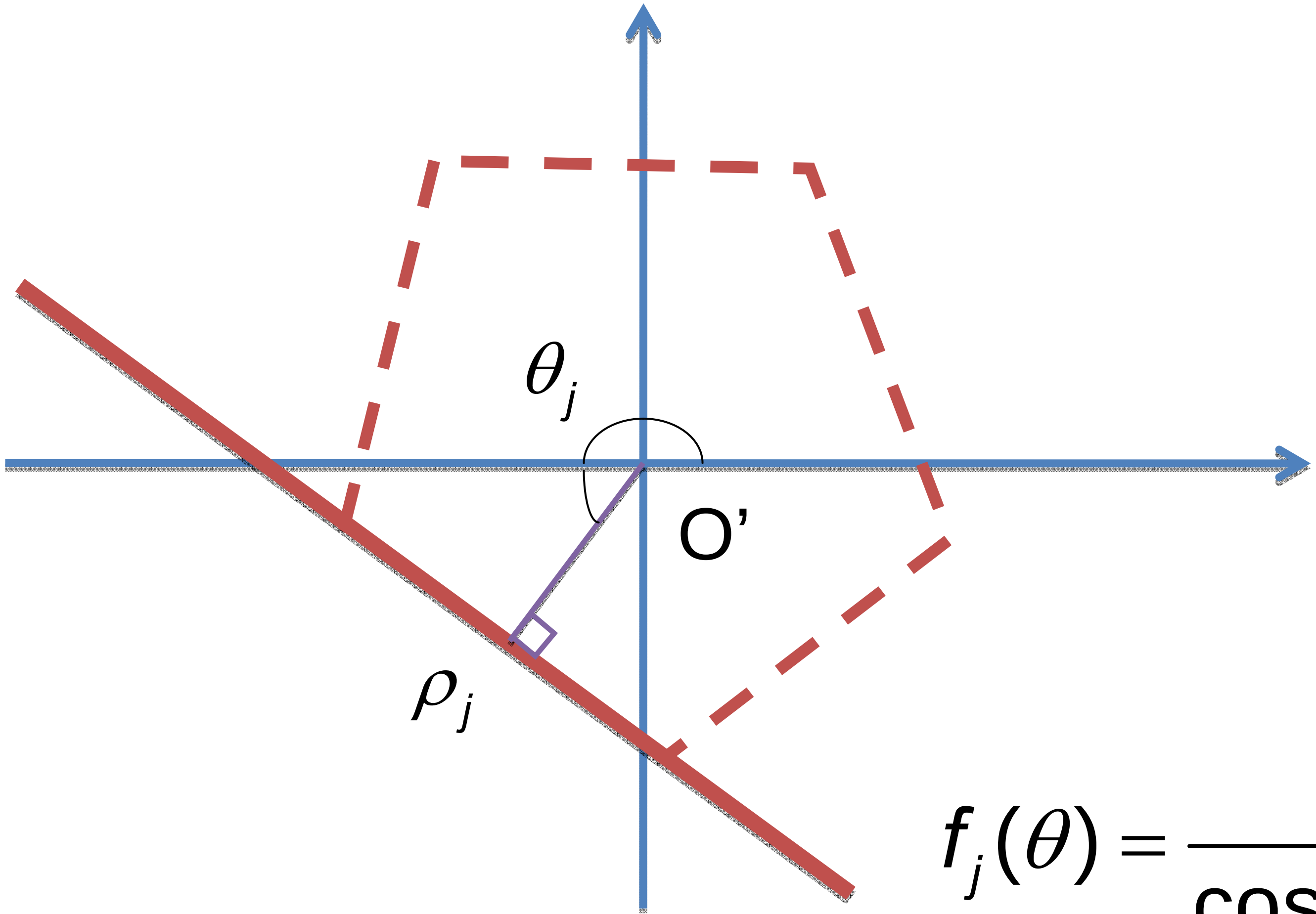
где

$$t_+ \equiv \begin{cases} t, & t \geq 0 \\ 0 & t < 0 \end{cases} = t \cdot (t \geq 0)$$

Многоугольная RBF:

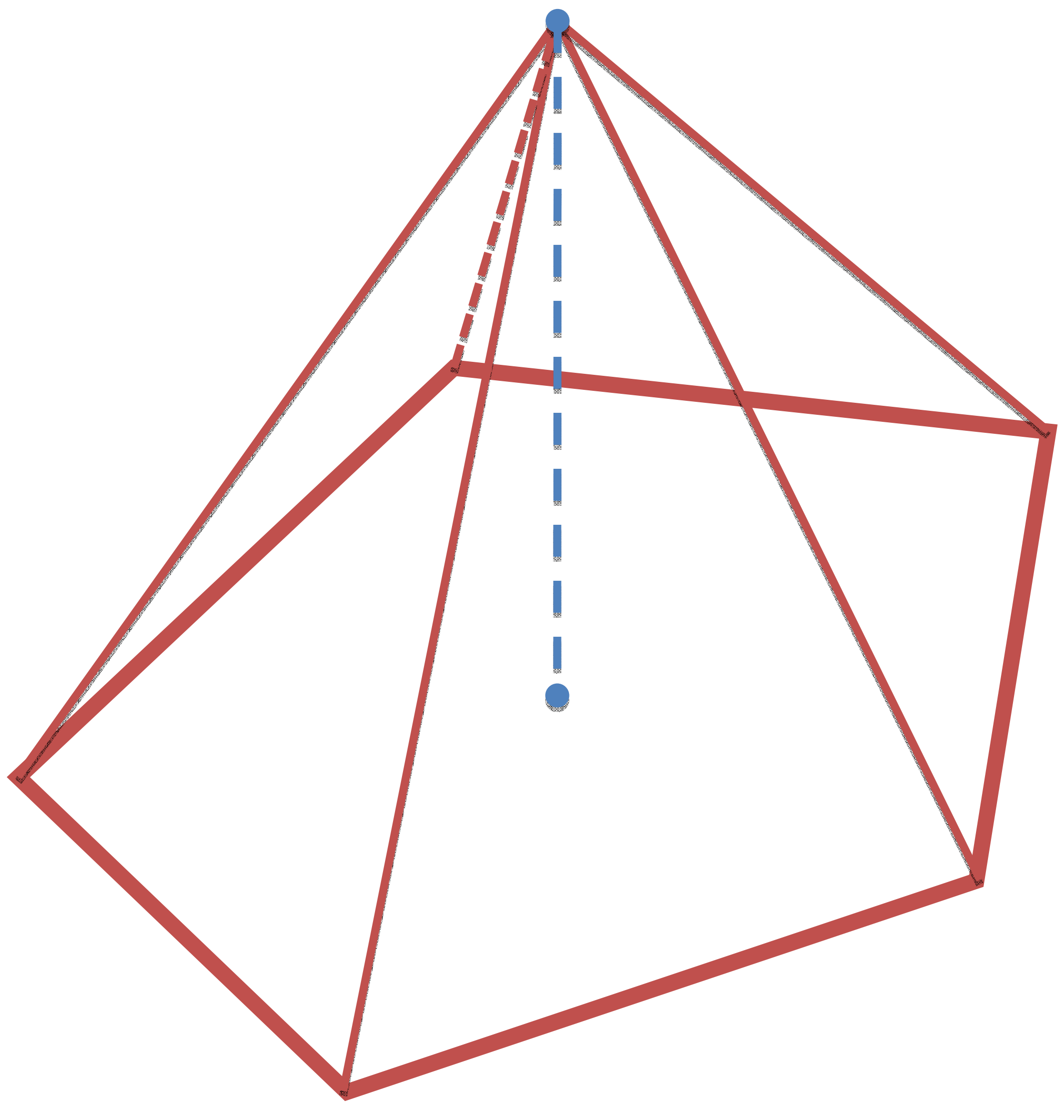


Граница получается из полярных уравнений прямых:



$$f_j(\theta) = \frac{\rho_j}{\cos(\theta - \theta_j)}$$

Функция-«шапочка»:



Дифференцируемость (гладкость) на границе:

$$\frac{\partial \varphi_j}{\partial \mathbf{n}} = 0$$

где n – внешняя нормаль к границе.

$$\varphi = \left(1 - \frac{\rho}{f(\theta)}\right)_+^2$$

RBF с гладкой вершиной:

$$\varphi = \left(1 + \frac{2\rho}{f(\theta)}\right)\left(1 - \frac{\rho}{f(\theta)}\right)_+^2$$

Многомерные RBF:

$$\varphi_j = \varphi_j(\rho, \mathbf{V})$$

где

$$\rho = \left\| \mathbf{x} - \mathbf{c}_j \right\|$$

$\mathbf{V}$ – вектор на единичной гиперсфере

Алгебрологический синтез RBF

Пусть Ω_1 и Ω_2 – звездные относительно начала координат (полюса) области, с границами, описываемыми уравнениями

$$\rho = f_1(\theta) \quad \text{на } \partial\Omega_1$$

$$\rho = f_2(\theta) \quad \text{на } \partial\Omega_2$$

При этом

$$\begin{cases} f_1(\theta) - \rho > 0 & в\ \Omega_1 \\ f_1(\theta) - \rho < 0 & в\ R \setminus \overline{\Omega}_1 \end{cases}$$

$$\begin{cases} f_2(\theta) - \rho > 0 & в\ \Omega_2 \\ f_2(\theta) - \rho < 0 & в\ R \setminus \overline{\Omega}_2 \end{cases}$$

Пусть Ω - сложная область (звездная):

$$\Omega = \Omega_1 \cap \Omega_2$$

или

$$\Omega = \Omega_1 \cup \Omega_2$$

тогда

$$\rho = \min\left\{f_1(\theta), f_2(\theta)\right\}$$

или

$$\rho = \max\left\{f_1(\theta), f_2(\theta)\right\}$$

Так как

$$\min\{x, y\} = \frac{1}{2}(x + y - |x - y|)$$

$$\max\{x, y\} = \frac{1}{2}(x + y + |x - y|)$$

то

$$\rho(\theta) = f_1 \wedge f_2 =$$

$$= \frac{1}{2}\left[f_1(\theta) + f_2(\theta) - |\, f_1(\theta) - f_2(\theta)\,|\right]$$

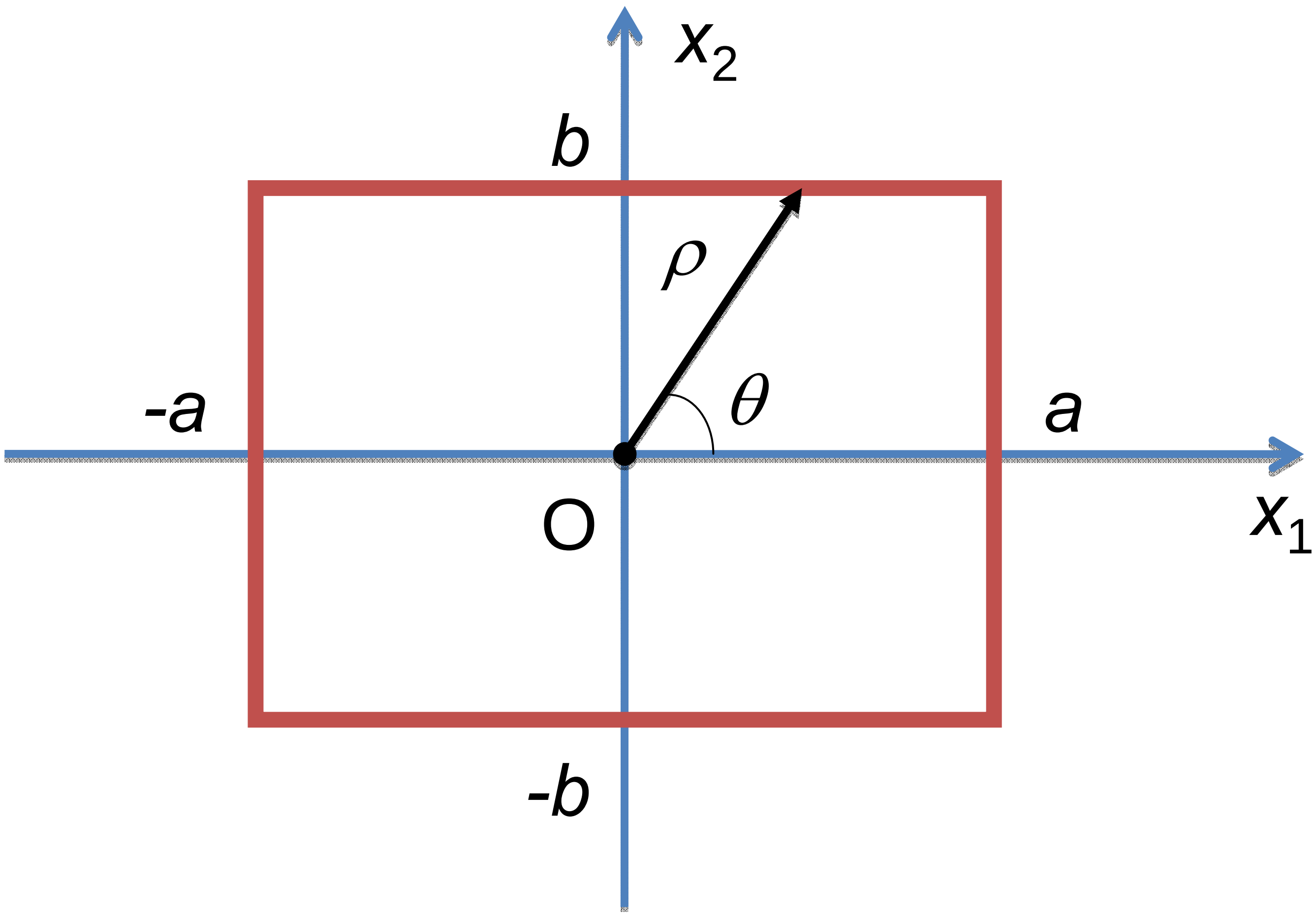
или

$$\rho(\theta) = f_1 \vee f_2 =$$

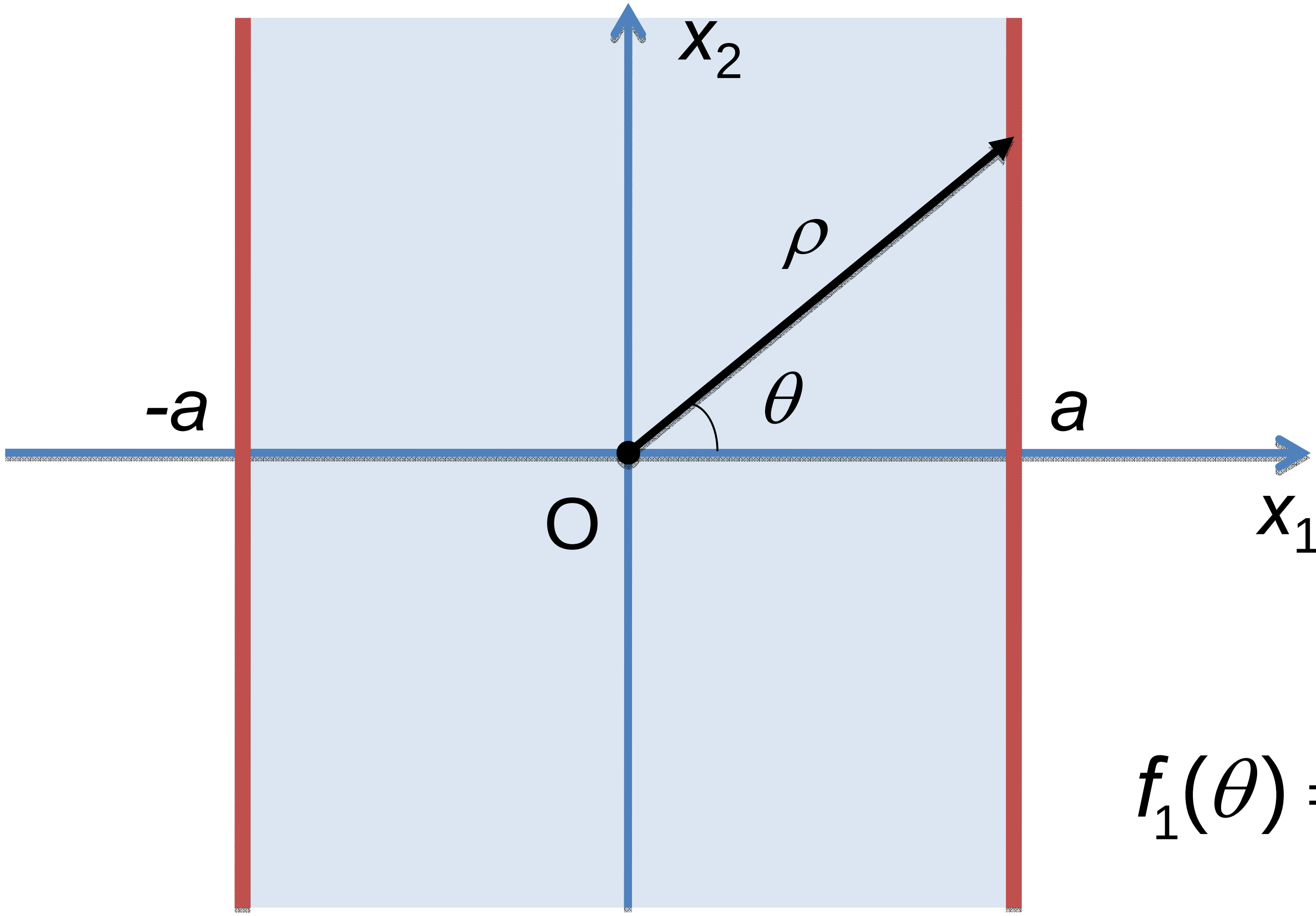
$$= \frac{1}{2}\left[f_1(\theta) + f_2(\theta) + |\, f_1(\theta) - f_2(\theta)\,|\right]$$

Пример 1. Прямоугольник

$$[-a, a] \times [-b, b]$$

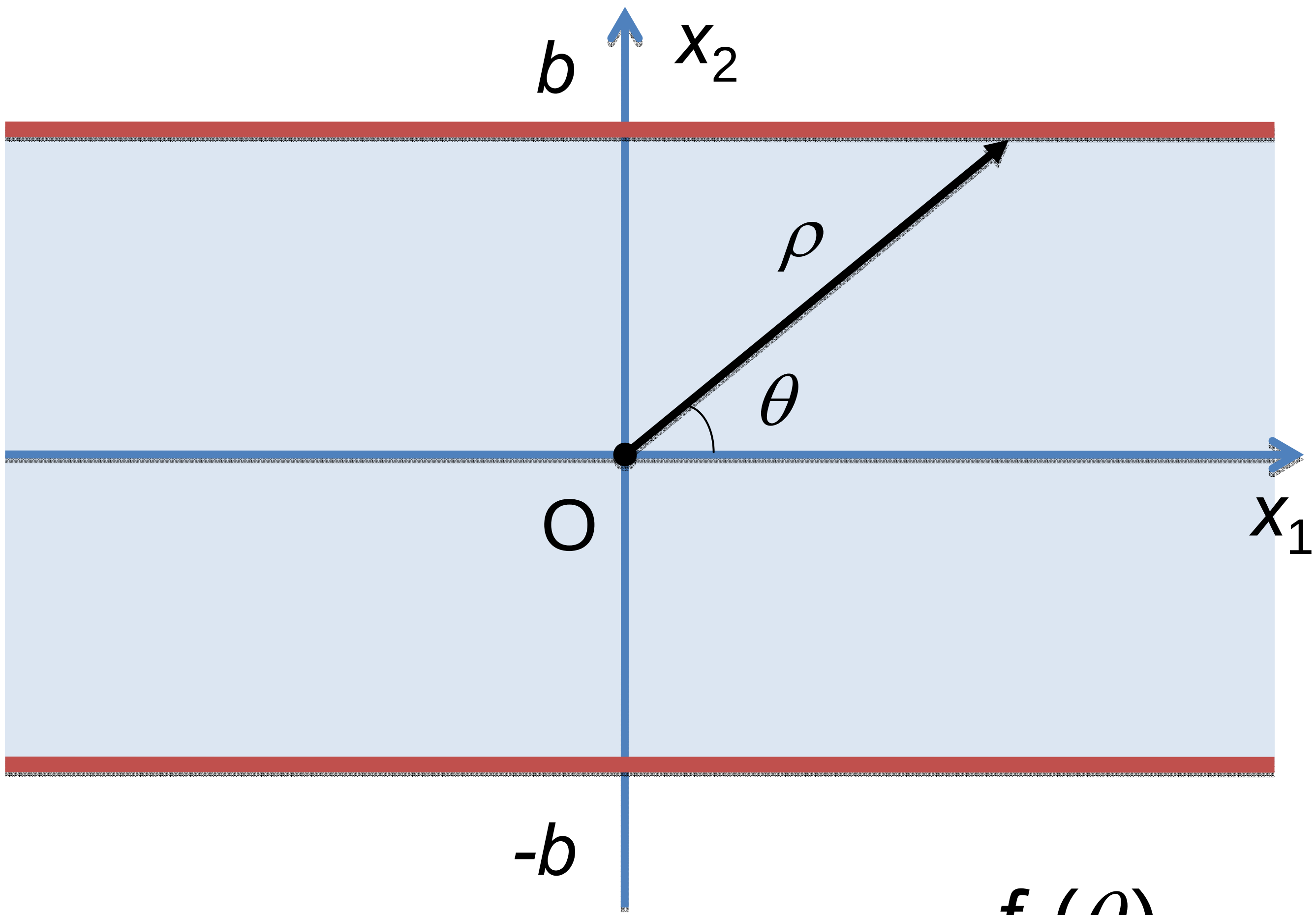


Полоса, параллельная оси **OY**



$$f_1(\theta) = \frac{a}{|\cos\theta|}$$

Полоса, параллельная оси **OX**



$$f_2(\theta) = \frac{b}{|\sin\theta|}$$

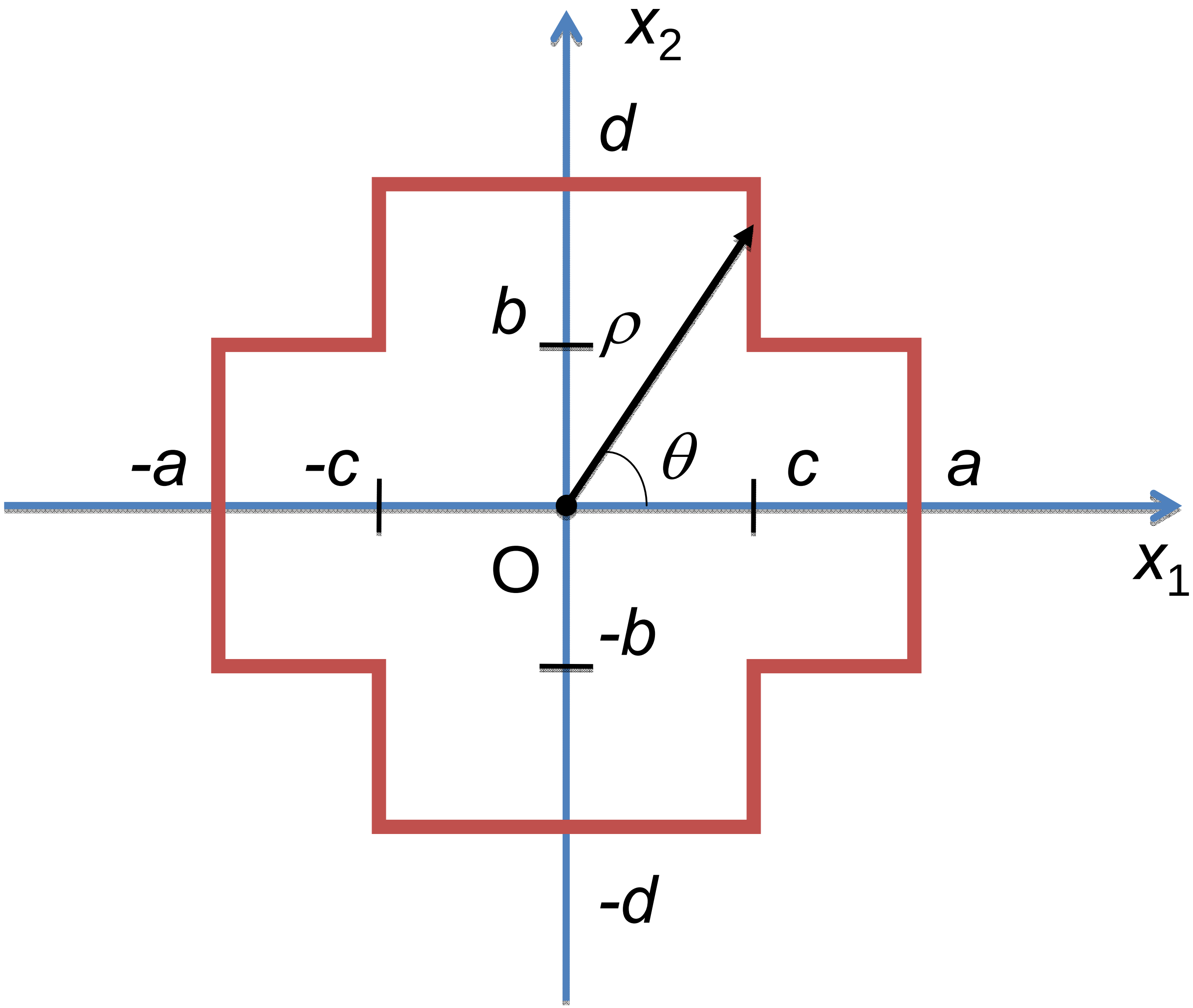
$$\rho(\theta) = f_1(\theta) \wedge f_2(\theta)$$

$$f_1(\theta) = \frac{a}{|\cos\theta|}, \quad f_2(\theta) = \frac{b}{|\sin\theta|}$$

$$\rho(\theta) =$$

$$= \frac{1}{2}\left[\frac{a}{|\cos\theta|} + \frac{b}{|\sin\theta|} - \left|\frac{a}{|\cos\theta|} - \frac{b}{|\sin\theta|}\right|\right]$$

Пример 2. Крест



$$f_1(\theta) = \frac{a}{|\cos\theta|}, \quad f_2(\theta) = \frac{b}{|\sin\theta|}$$

$$f_3(\theta) = \frac{c}{|\cos\theta|}, \quad f_4(\theta) = \frac{d}{|\sin\theta|}$$

$$\rho(\theta) = \left[f_1(\theta) \wedge f_2(\theta)\right] \vee \left[f_3(\theta) \wedge f_4(\theta)\right]$$

Сглаживание углов:

$$|x| \to \sqrt{x^2 + \varepsilon} \qquad 0 < \varepsilon << 1$$

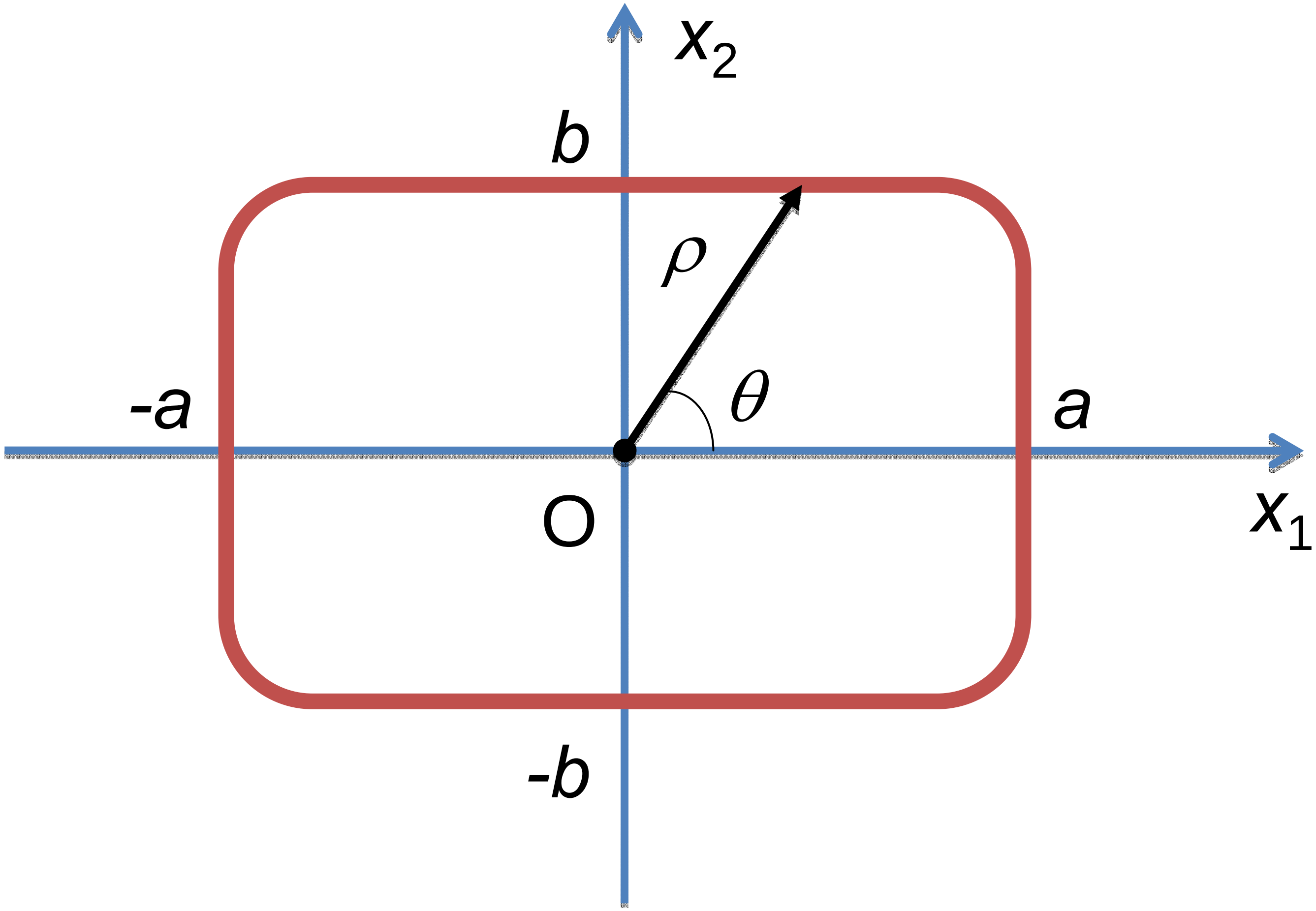
$$\rho(\theta) = f_1 \wedge f_2 =$$

$$= \frac{1}{2}\left[f_1(\theta) + f_2(\theta) - \sqrt{\left[f_1(\theta) - f_2(\theta)\right]^2 + \varepsilon} \right]$$

или

$$\rho(\theta) = f_1 \vee f_2 =$$

$$= \frac{1}{2}\left[f_1(\theta) + f_2(\theta) + \sqrt{\left[f_1(\theta) - f_2(\theta)\right]^2 + \varepsilon} \right]$$



$$\rho(\theta) = \frac{1}{2}\left[\frac{a}{|\cos\theta|} + \frac{b}{|\sin\theta|} - \sqrt{\left(\frac{a}{|\cos\theta|} - \frac{b}{|\sin\theta|}\right)^2 + \varepsilon}\right]$$