UNIVERSITY OF LEEDS

# Class: Machine Learning

## Support Vector Machines – part 3

Instructor: Matteo Leonetti

# Learning outcomes

UNIVERSITY OF LEEDS

- Derive the dual formulation of a constrained optimisation problem

# Substitution

minimise: $$\frac{1}{2}\|\boldsymbol{w}\|^2$$

Subject to the constraints: $$t_i\left(\boldsymbol{w}^{\boldsymbol{T}}\Phi\left(\boldsymbol{x}_i\right)+w_0\right)\geq 1$$

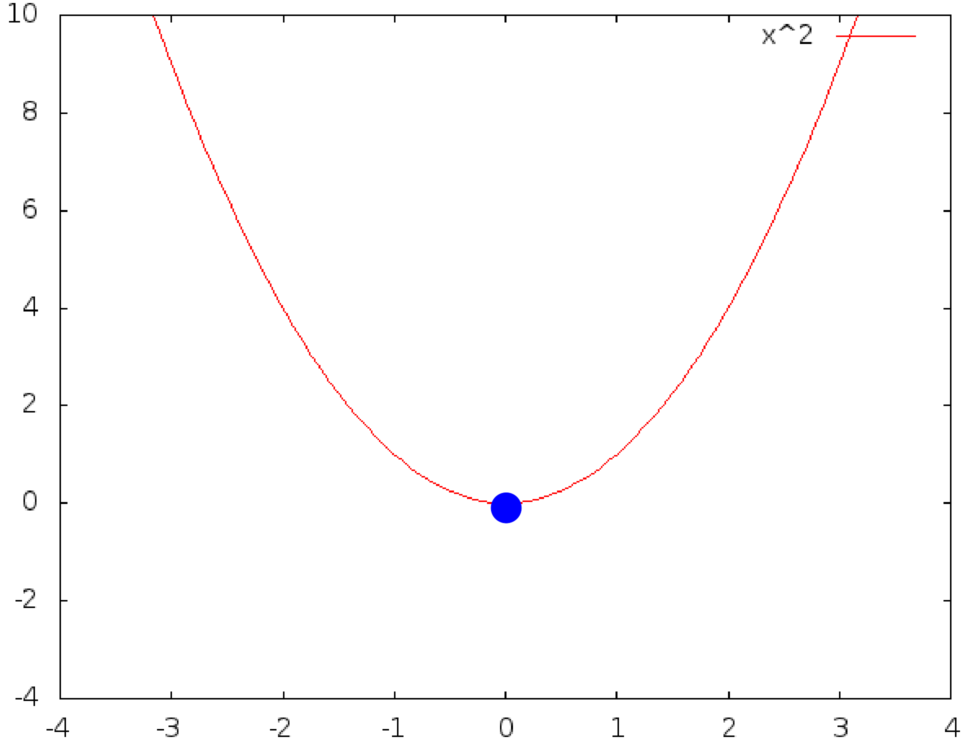
This way we would have a higher dimensional problem, which is also more difficult to solve.

Is there a better formulation?

# Duality Theory

# Example - unconstrained

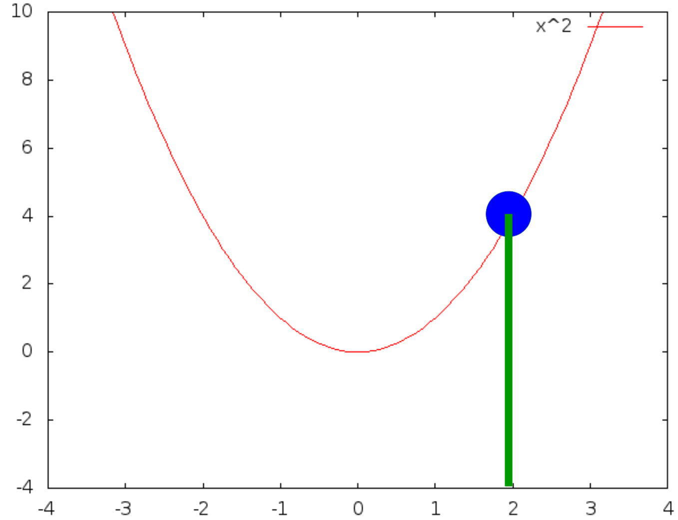
UNIVERSITY OF LEEDS




$$min\, x^2$$

$$\nabla_x f(x) = 2x = 0 \quad \Rightarrow x = 0$$

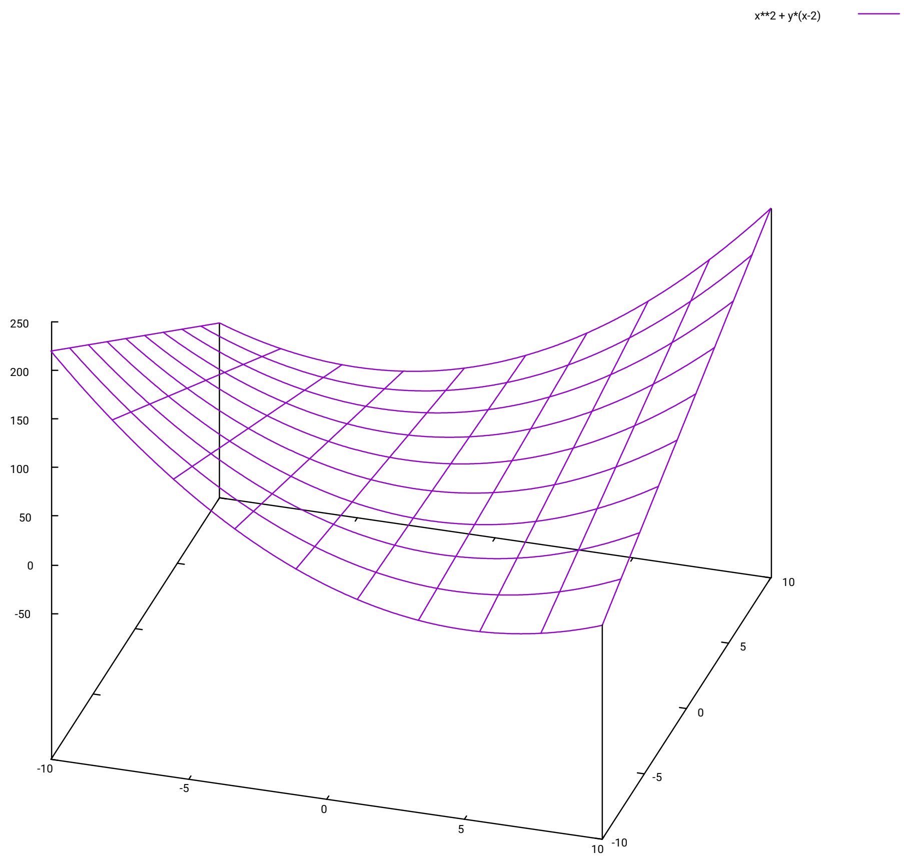
# Example - constrained




$$\min \; x^2$$

$$\text{s.t.} \quad x = 2$$

# The Lagrangian



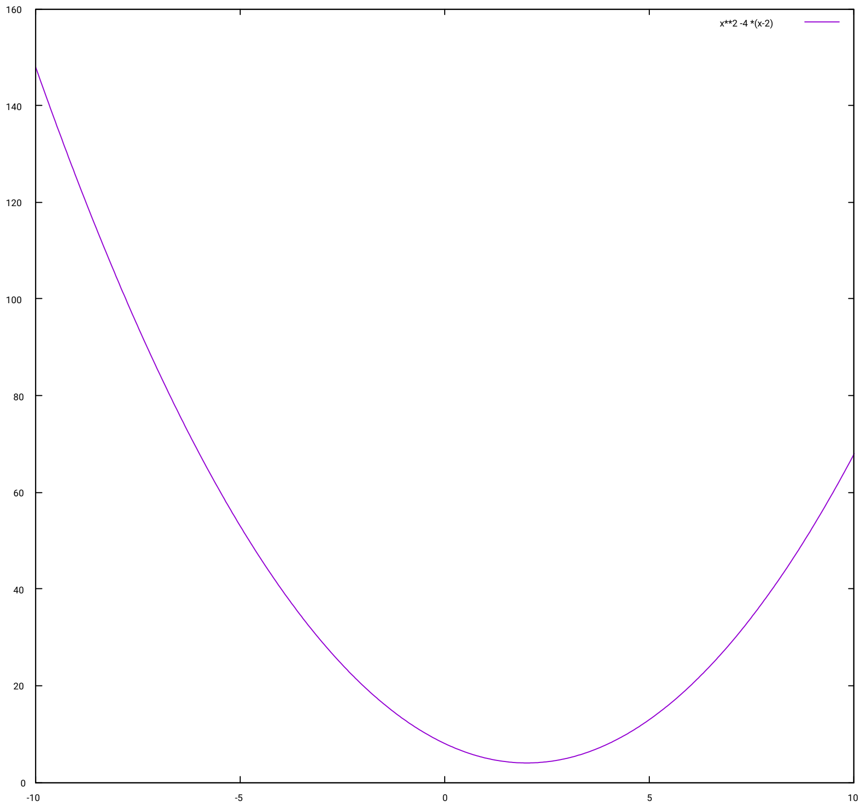

$$\min_x \max_\lambda \; x^2 + \lambda(x-2)$$

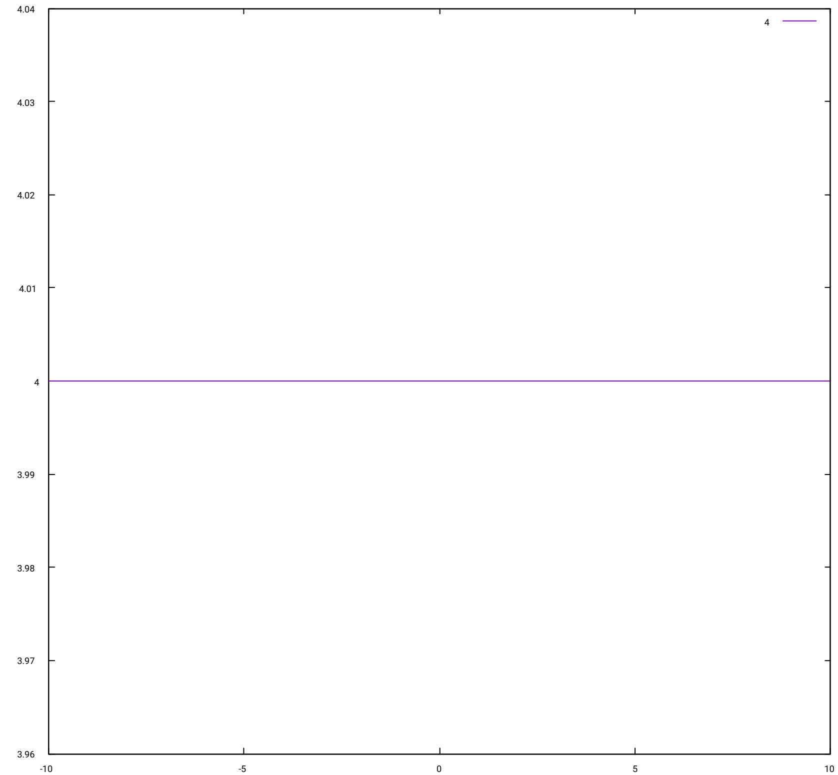
$$\begin{cases} \nabla_x f(x,\lambda) = 2x + \lambda = 0 \\ \nabla_\lambda f(x,\lambda) = x - 2 = 0 \end{cases}$$

$$\boxed{\begin{aligned} x &= 2 \\ \lambda &= -4 \end{aligned}}$$

$$L(x,\lambda) = x^2 + \lambda(x-2)$$

# The Lagrangian - sliced

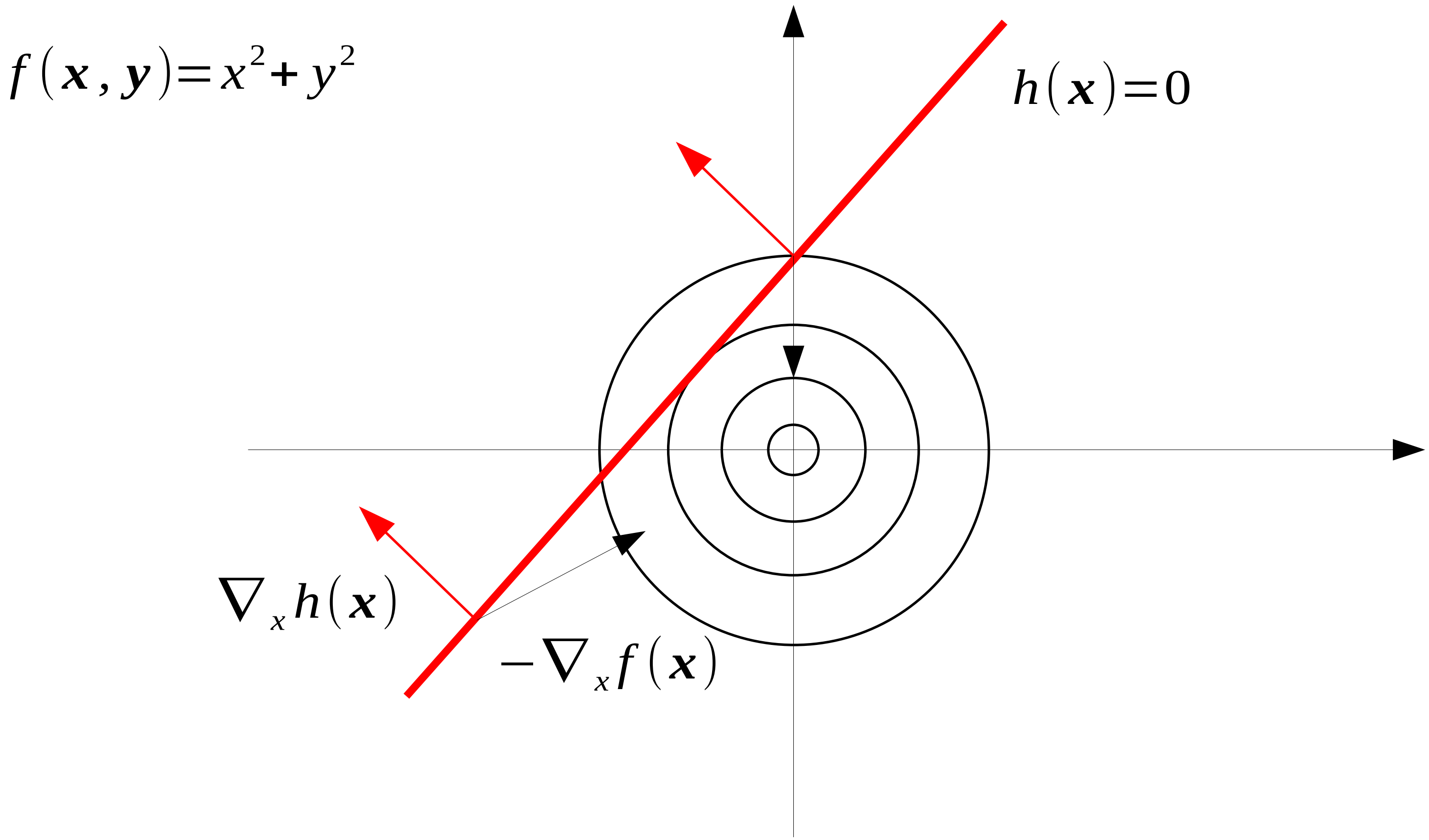



$$L(x,-4)=x^2-4(x-2)$$

$$L(2,\lambda)=4$$

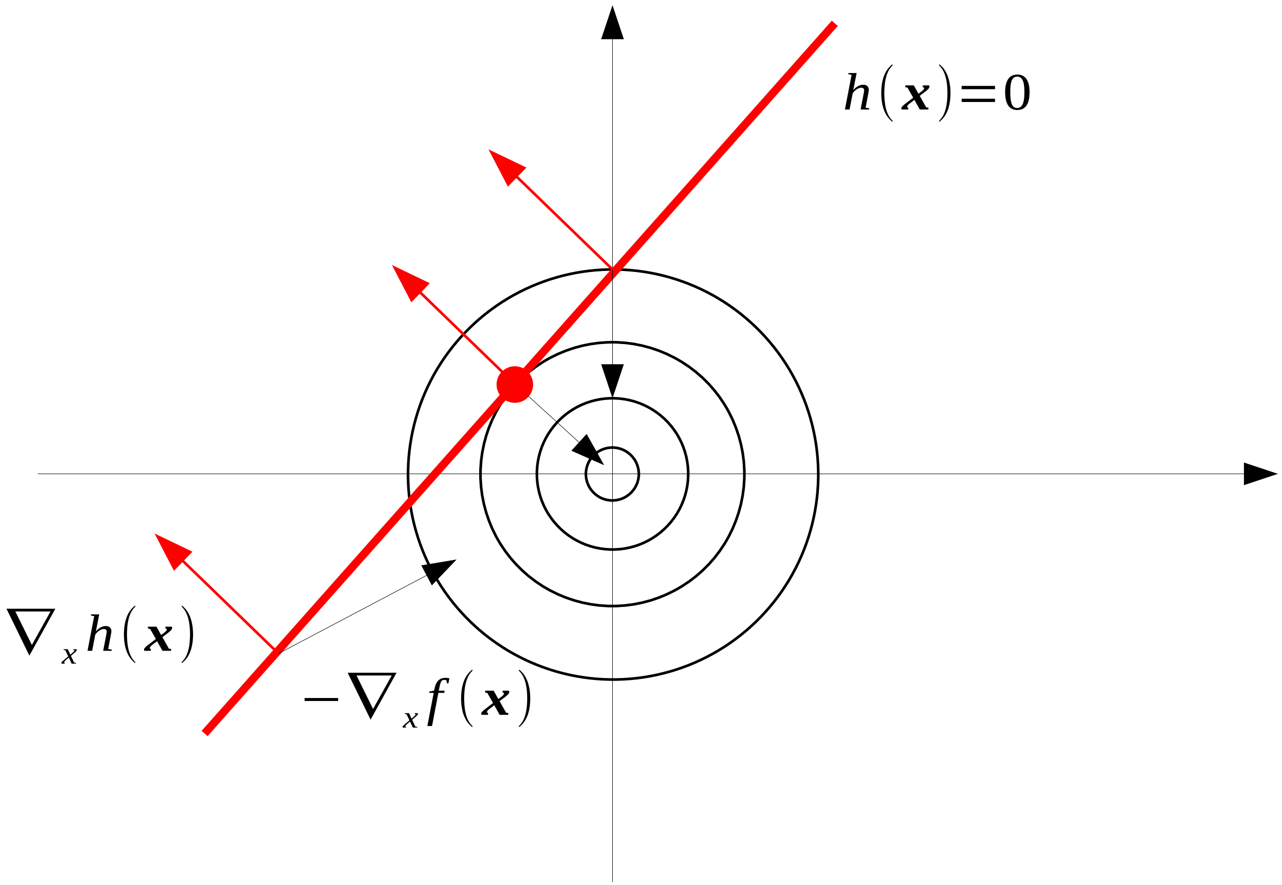
# Lagrange Multipliers




By looking at the gradients, can you tell when a point is a local minimum for the constrained problem?

# Lagrange Multipliers

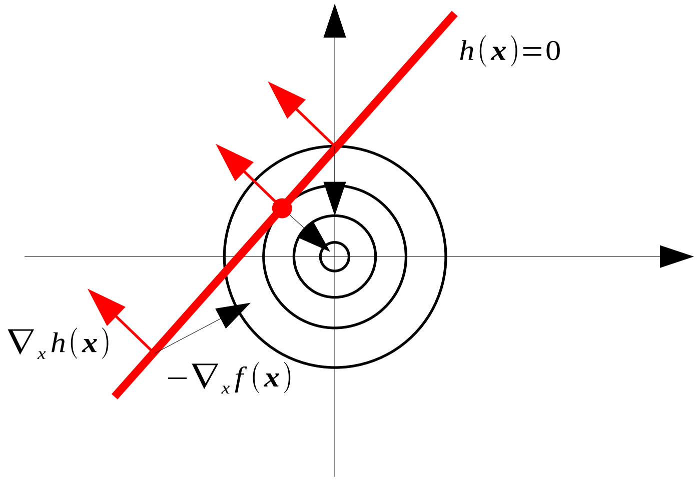
UNIVERSITY OF LEEDS




When the gradients are parallel! $-\nabla_x f(\boldsymbol{x})=\lambda \nabla_x h(\boldsymbol{x})$

# Lagrange Multipliers

UNIVERSITY OF LEEDS




When the gradients are parallel! $-\nabla_x f(\boldsymbol{x}) = \lambda \nabla_x h(\boldsymbol{x})$

This is achieved by: $\nabla_x L(\boldsymbol{x}, \boldsymbol{\lambda}) = 0$

since:

$$\nabla_x L(\boldsymbol{x}, \boldsymbol{\lambda}) = \nabla_x f(\boldsymbol{x}) + \lambda \nabla_x h(\boldsymbol{x}) = 0$$

# Lagrange multipliers

UNIVERSITY OF LEEDS

$$\min \quad f(\boldsymbol{x})$$

Subject to

$$h_i(\boldsymbol{x})=0 \quad \forall\, i=1,\ldots,m$$

It is possible to form a function such that its stationary points are optimal solutions to the original problem:

$$L(\boldsymbol{x},\boldsymbol{\lambda})=f(\boldsymbol{x})+\sum_{i=1}^{m}\lambda_i h_i(\boldsymbol{x})$$

$\nabla_x L(\boldsymbol{x},\boldsymbol{\lambda})=\nabla_x f(\boldsymbol{x})+\lambda\nabla_x h(\boldsymbol{x})=0$ Ensures that the gradients are parallel

$\nabla_\lambda L(\boldsymbol{x},\boldsymbol{\lambda})=h(\boldsymbol{x})=0$ Ensures that the solution satisfies the constraints

# Question

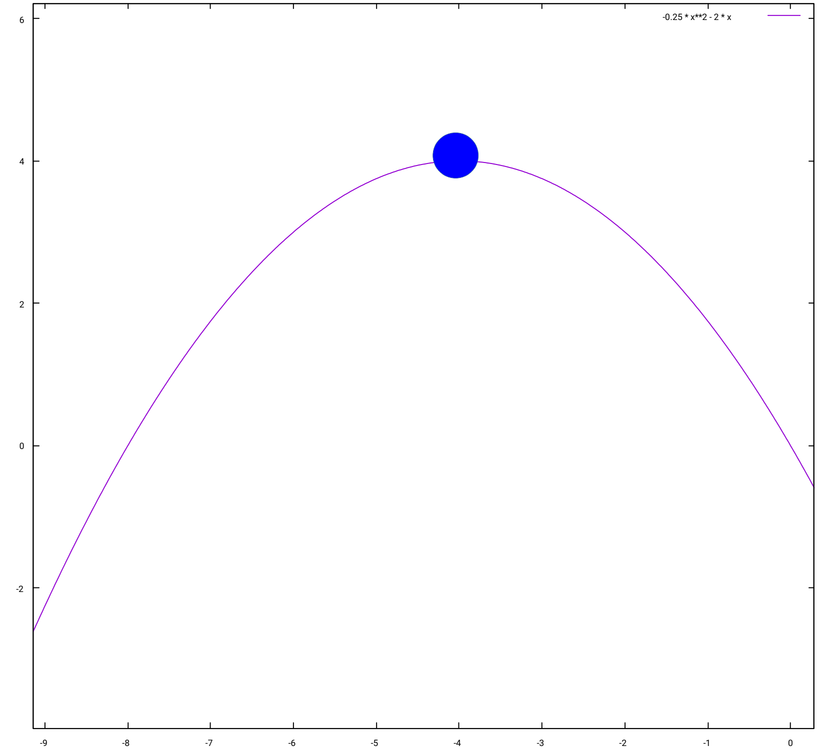
How many variables lambda did I add?

# The Dual Problem

UNIVERSITY OF LEEDS

$$L(x,\lambda)=x^2+\lambda(x-2)$$

$$\nabla_x f(x,\lambda)=2x+\lambda=0 \qquad x=-\frac{1}{2}\lambda$$



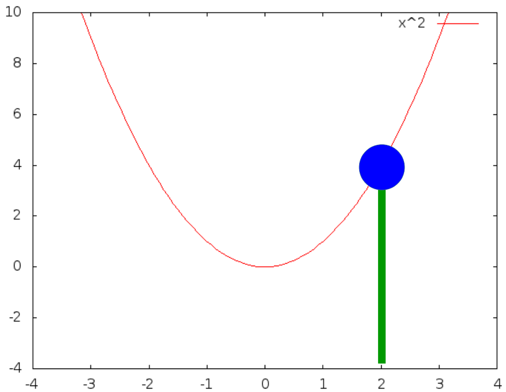

Substitute x:

$$q(\lambda)=\left(-\frac{1}{2}\lambda\right)^2+\lambda\left(-\frac{1}{2}\lambda-2\right)=\frac{1}{4}\lambda^2-\frac{1}{2}\lambda^2-2\lambda$$

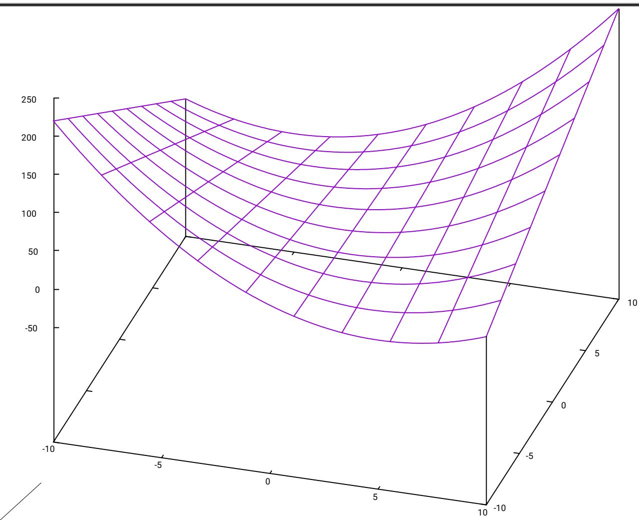
$$=-\frac{1}{4}\lambda^2-2\lambda$$

$$\nabla_\lambda q=-\frac{1}{2}\lambda-2=0 \qquad \lambda=-4$$

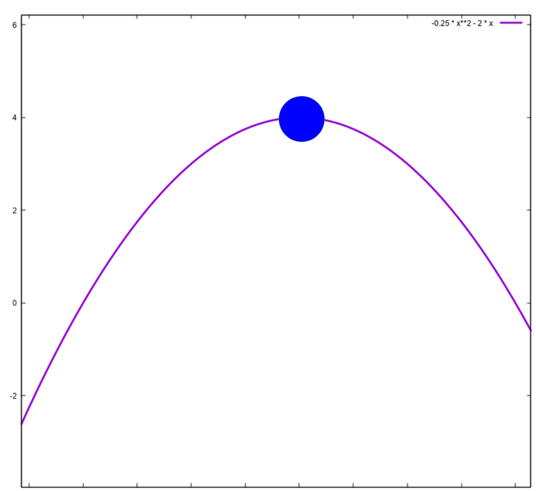
# The Dual Problem




$$\min \; f(x) = x^2$$
$$\text{s.t.} \quad x = 2$$



$$L(x, \lambda) = x^2 + \lambda(x - 2)$$

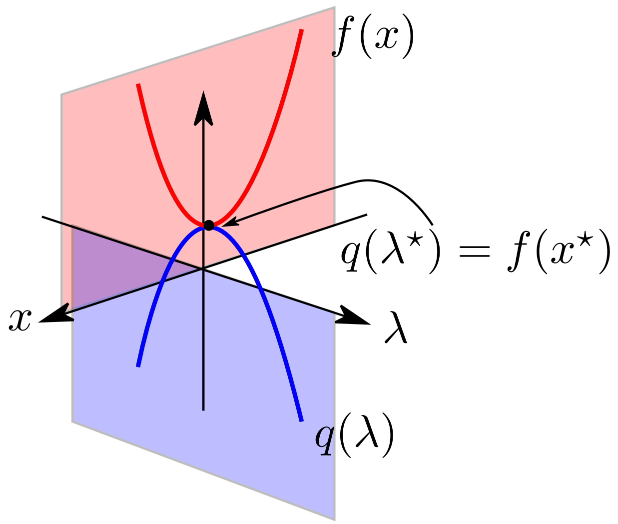



$$\max \; q(\lambda) = -\frac{1}{4}\lambda^2 - 2\lambda$$

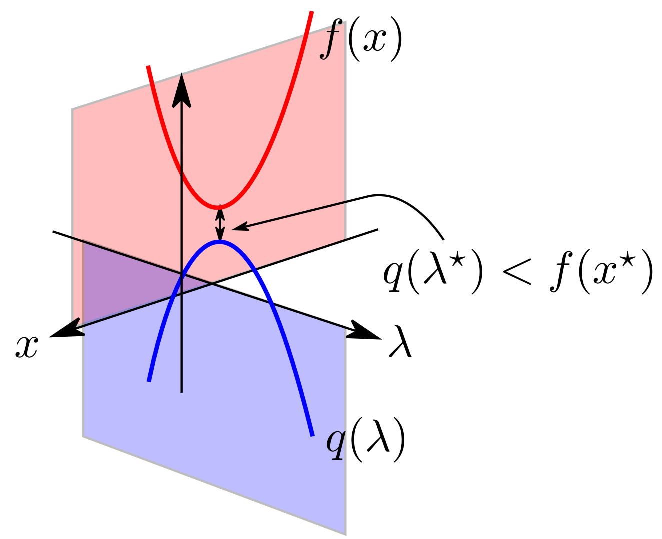
$$\boxed{f(2) = q(-4) = 4}$$

# Duality



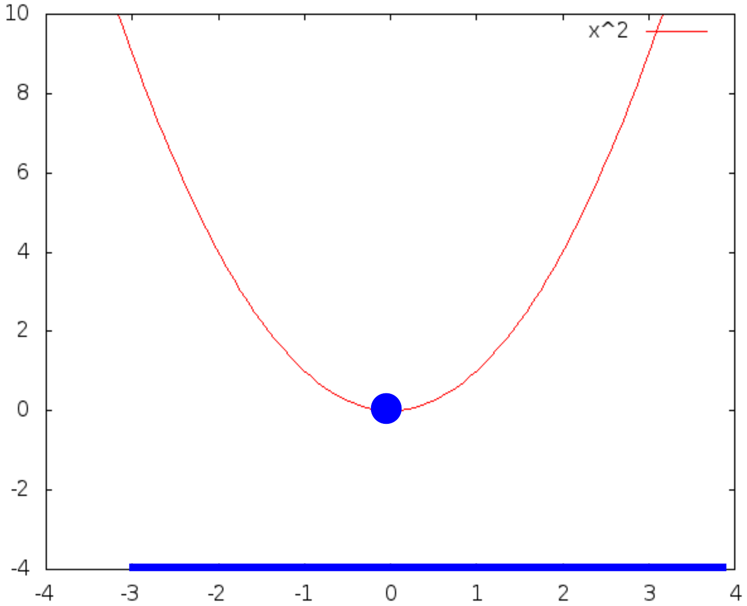



strong duality




weak duality

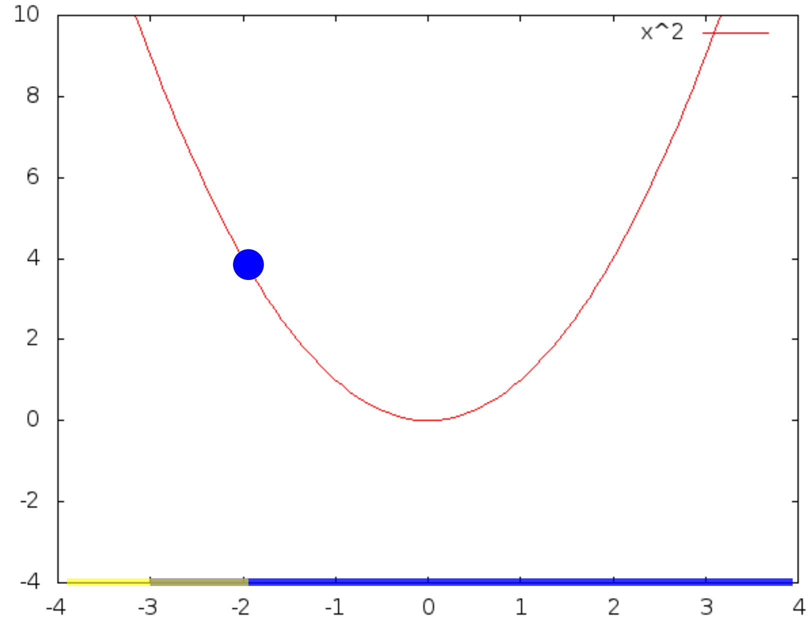
# Inequality Constraints




$$\min\ f(x)=x^2$$

s.t. $x \geq -3$

Minimum inside the constraint




$$\min\ f(x)=x^2$$

s.t. $x \geq -3$

$x \leq -2$

Minimum on the border

# Karush–Kuhn–Tucker conditions

UNIVERSITY OF LEEDS

Extend lagrangian multipliers to inequality constraints

$$\min f(\vec{x})$$

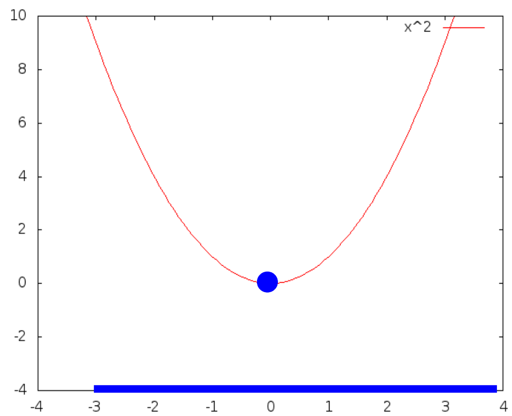
Subject to

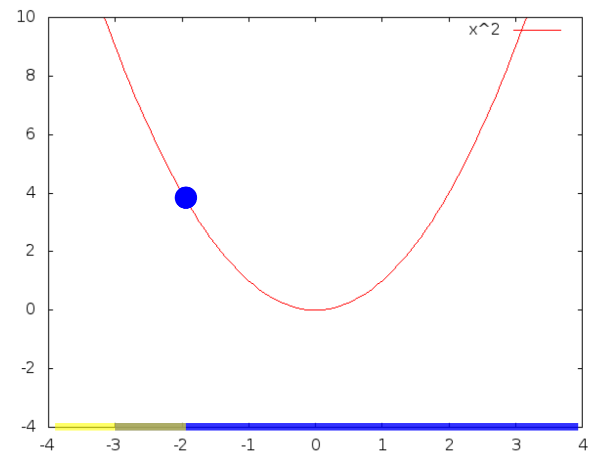
$$h_i(\vec{x}) \leq 0 \quad \forall i = 1, \ldots, n$$

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i h_i(\boldsymbol{x})$$

$\nabla_x L(\boldsymbol{x}, \boldsymbol{\lambda}) = 0$ What else?

# Complementary Slackness




$$\min \quad L(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i h_i(\boldsymbol{x})$$

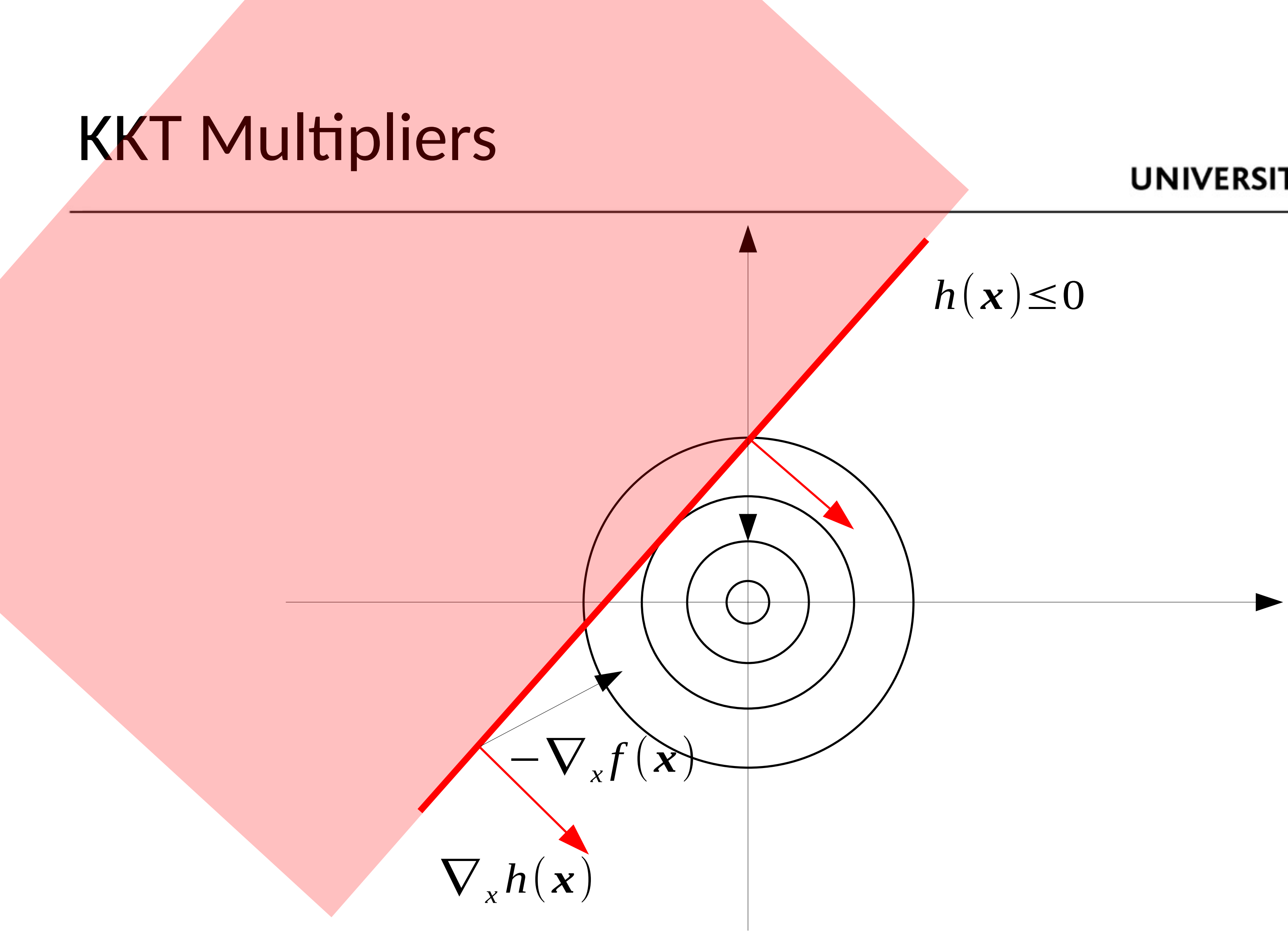
$$\text{s.t.} \quad h_i(\boldsymbol{x}) \leq 0$$

For Inactive constraints:

For active constraints

$$\lambda_j = 0 \longrightarrow \lambda_j h_j(\boldsymbol{x}) = 0 \longleftarrow h_j(\boldsymbol{x}) = 0$$

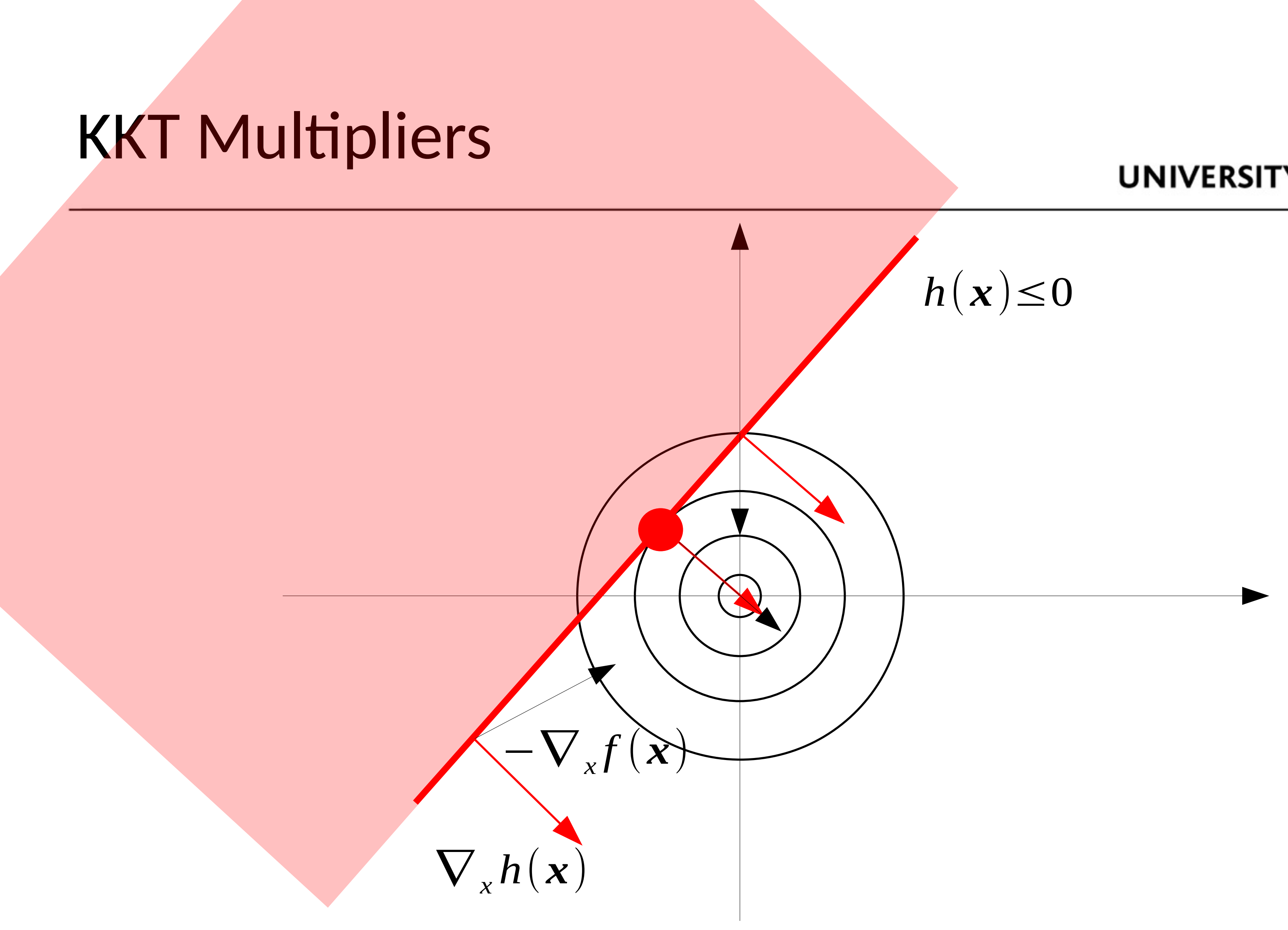
# KKT Multipliers




With inequality constraints, not only the gradients must be parallel, but also?

# KKT Multipliers

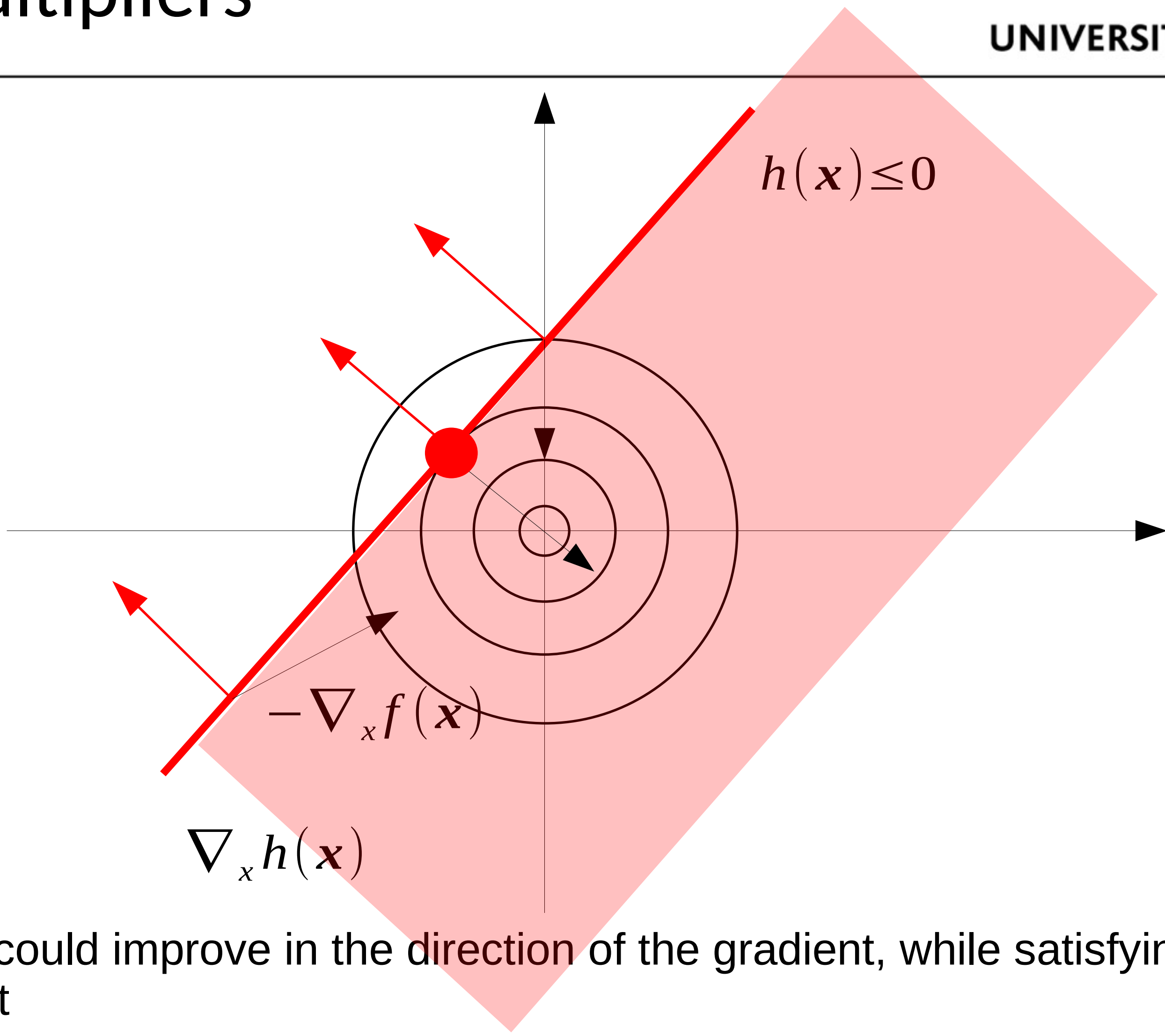





With inequality constraints, not only the gradients must be parallel, but the antigradient must have the same direction as the gradient of the constraint!

$$-\nabla_x f(\boldsymbol{x}) = \lambda \nabla_x h(\boldsymbol{x}) \qquad \lambda \geq 0$$

# KKT Multipliers




Otherwise, I could improve in the direction of the gradient, while satisfying the constraint

$$-\nabla_x f(\boldsymbol{x}) = \lambda \nabla_x h(\boldsymbol{x}) \qquad \lambda \geq 0$$

# KKT Conditions



min $f(\boldsymbol{x})$

min $f(\boldsymbol{x})$

Subject to

$$h_i(\boldsymbol{x})=0 \quad \forall i=1,\ldots,m$$

min $f(\boldsymbol{x})$

Subject to

$$h_i(\boldsymbol{x})\leq 0 \quad \forall i=1,\ldots,m$$

Corresponding system of equations

$$\nabla_x f(\boldsymbol{x})=0$$

$$\nabla_x L(\boldsymbol{x},\boldsymbol{\lambda})=0$$

$$\nabla_\lambda L(\boldsymbol{x},\boldsymbol{\lambda})=0$$

$$\nabla_x L(\boldsymbol{x},\boldsymbol{\lambda})=0$$

$$\lambda_i g_i(\boldsymbol{x})=0 \quad \forall i=1,\ldots,n$$

$$\lambda_i\geq 0 \quad \forall i=1,\ldots,n$$

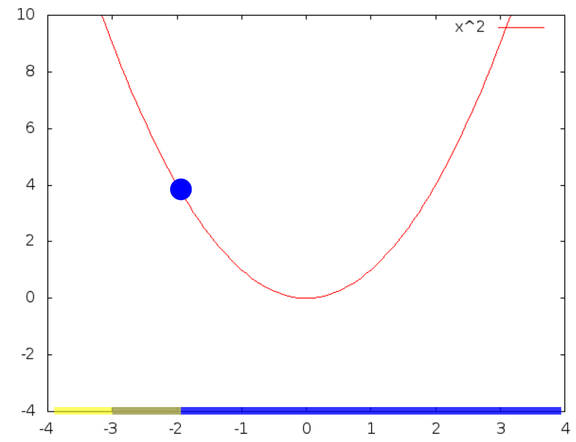
# Example

Lagrangian: $$L(x,\boldsymbol{\lambda}) = x^2 + \lambda_1(-x-3) + \lambda_2(x+2)$$

s.t. $$\lambda_1, \lambda_2 \geq 0$$

$$\lambda_1(-x-3) = 0$$

$$\lambda_2(x+2) = 0$$

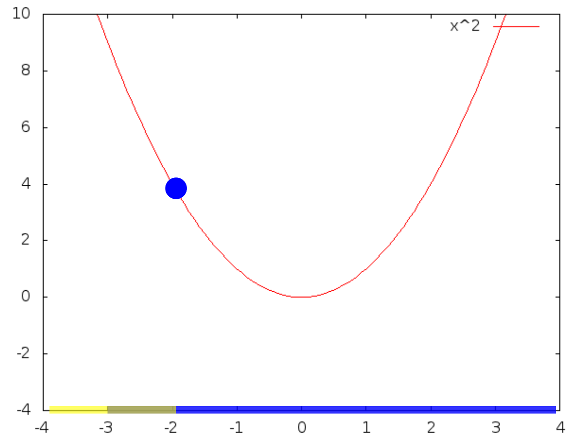


# Stationary point for the Lagrangian

UNIVERSITY OF LEEDS

Lagrangian: $L(x,\boldsymbol{\lambda}) = x^2 + \lambda_1(-x-3) + \lambda_2(x+2)$

s.t. $\boxed{\lambda_1, \lambda_2 \geq 0}$




$$\begin{cases} \nabla_x L(x,\boldsymbol{\lambda}) = 2x - \lambda_1 + \lambda_2 = 0 \\ \lambda_1(-x-3) = 0 \\ \lambda_2(x+2) = 0 \end{cases}$$

Let's assume that the first constraint, -x -3, is active and x = -3

$$x = -3$$

$$\downarrow$$

$$\lambda_2(-3+2) = 0 \quad \lambda_2 = 0$$

$$\downarrow$$

$$-6 - \lambda_1 + 0 = 0 \quad \boxed{\lambda_1 = -6}$$

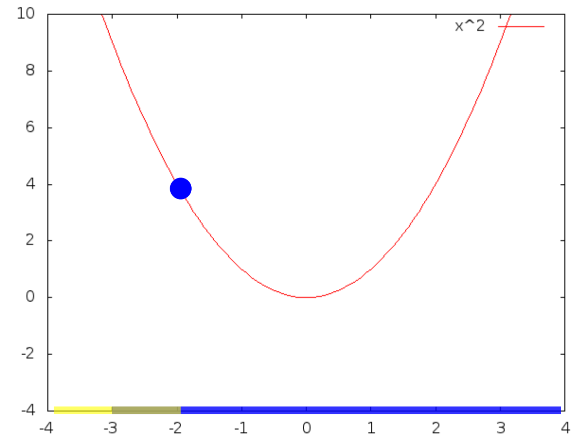
It would violate the constraint on $\lambda_1$

# Stationary point for the Lagrangian

UNIVERSITY OF LEEDS

Lagrangian: $L(x,\boldsymbol{\lambda}) = x^2 + \lambda_1(-x-3) + \lambda_2(x+2)$

s.t. $\boxed{\lambda_1, \lambda_2 \geq 0}$ $\quad x+2 \leq 0$, $\quad -x-3 \leq 0$




$$\begin{cases} \nabla_x L(x,\boldsymbol{\lambda}) = 2x - \lambda_1 + \lambda_2 = 0 \\ \lambda_1(-x-3) = 0 \\ \lambda_2(x+2) = 0 \end{cases}$$

Let's now assume that the second constraint is active and x = -2

$$x = -2$$

$$\downarrow$$

$$\lambda_1(2-3) = 0 \qquad \lambda_1 = 0$$

$$\downarrow$$

$$-4 - 0 + \lambda_2 = 0 \qquad \boxed{\lambda_2 = 4}$$

OK!

$$\boxed{\begin{aligned} x &= -2 \\ \lambda_1 &= 0 \\ \lambda_2 &= 4 \end{aligned}}$$

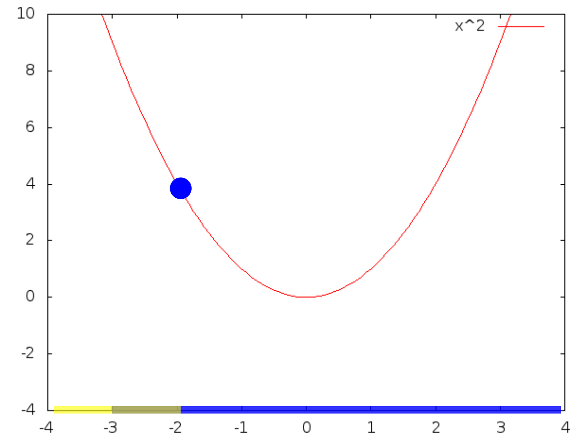
This is a stationary point of the Lagrangian AND the solution of the original constrained problem

# Dual Problem

UNIVERSITY OF LEEDS

Lagrangian: $L(x,\boldsymbol{\lambda}) = x^2 + \lambda_1(-x-3) + \lambda_2(x+2)$

s.t. $\boxed{\lambda_1, \lambda_2 \geq 0}$




Let's build the dual formulation!

$$\nabla_x L(x,\boldsymbol{\lambda}) = 2x - \lambda_1 + \lambda_2 = 0 \qquad x = \frac{\lambda_1 - \lambda_2}{2}$$

$$q(\boldsymbol{\lambda}) = \left(\frac{\lambda_1-\lambda_2}{2}\right)^2 + \lambda_1\left(-\frac{\lambda_1-\lambda_2}{2} - 3\right) + \lambda_2\left(\frac{\lambda_1-\lambda_2}{2} + 2\right)$$

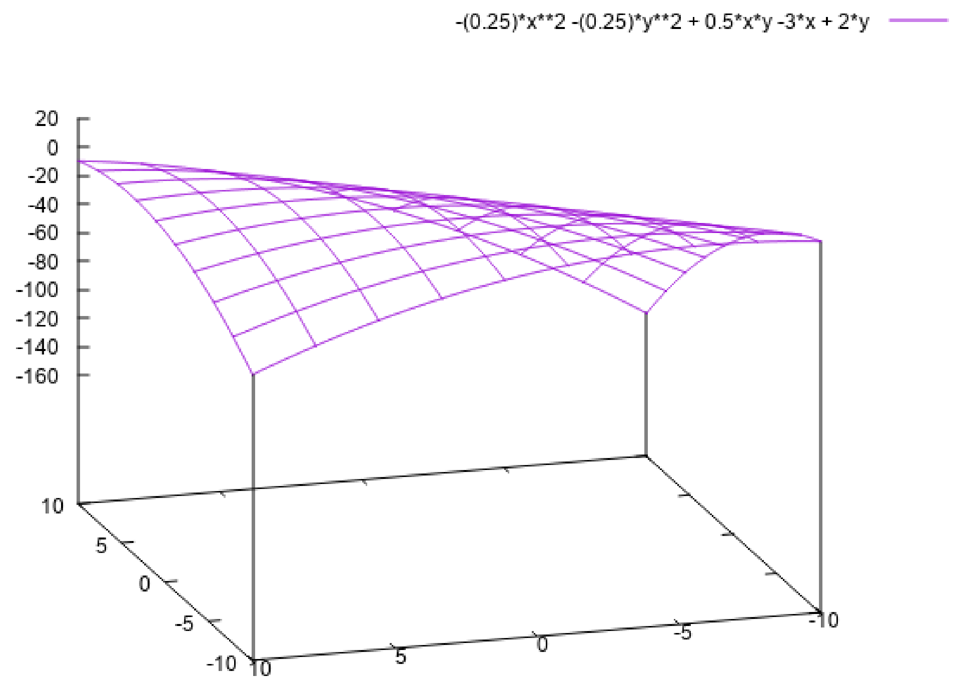
$$= \frac{1}{4}(\lambda_1^2 + \lambda_2^2 - 2\lambda_1\lambda_2) - \frac{1}{2}\lambda_1^2 + \frac{1}{2}\lambda_1\lambda_2 - 3\lambda_1 - \frac{1}{2}\lambda_2^2 + \frac{1}{2}\lambda_1\lambda_2 + 2\lambda_2$$

$$= -\frac{1}{4}\lambda_1^2 - \frac{1}{4}\lambda_2^2 - 3\lambda_1 + 2\lambda_2 + \frac{1}{2}\lambda_1\lambda_2$$

# Dual Problem

$$q(\boldsymbol{\lambda}) = -\frac{1}{4}\lambda_1^2 - \frac{1}{4}\lambda_2^2 - 3\lambda_1 + 2\lambda_2 + \frac{1}{2}\lambda_1\lambda_2$$

s.t.: $\lambda_1, \lambda_2 \geq 0$



# Question

UNIVERSITY OF LEEDS

Why did we do all this, again?

# Substitution



What is the dual formulation of this?

minimise: $\frac{1}{2}\|\boldsymbol{w}\|^2$

Subject to the constraints: $t_i\left(\boldsymbol{w}^{\boldsymbol{T}}\Phi\left(\boldsymbol{x}_i\right)+w_0\right)\geq 1$