BITS Pilani presentation

BITS Pilani
Pilani Campus

innovate achieve lead

BITS Pilani
Pilani Campus

SS ZC416 Mathematical Foundations for Data Science
Optimization

Objectives

- Finding a root of a function using Newton's Method

- Optimization using
  - Newton’s Method
  - Steepest Descent Method
  - Lagrange’s Multipliers

Newton's Method

The Problems

For solving

- The equation $f(x) = 0$
- Minimize $f(x)$

The Newton's Method

- Is based on the geometry of a curve, using the tangent lines to a curve.
- It requires calculus, in particular differentiation

The Intuition

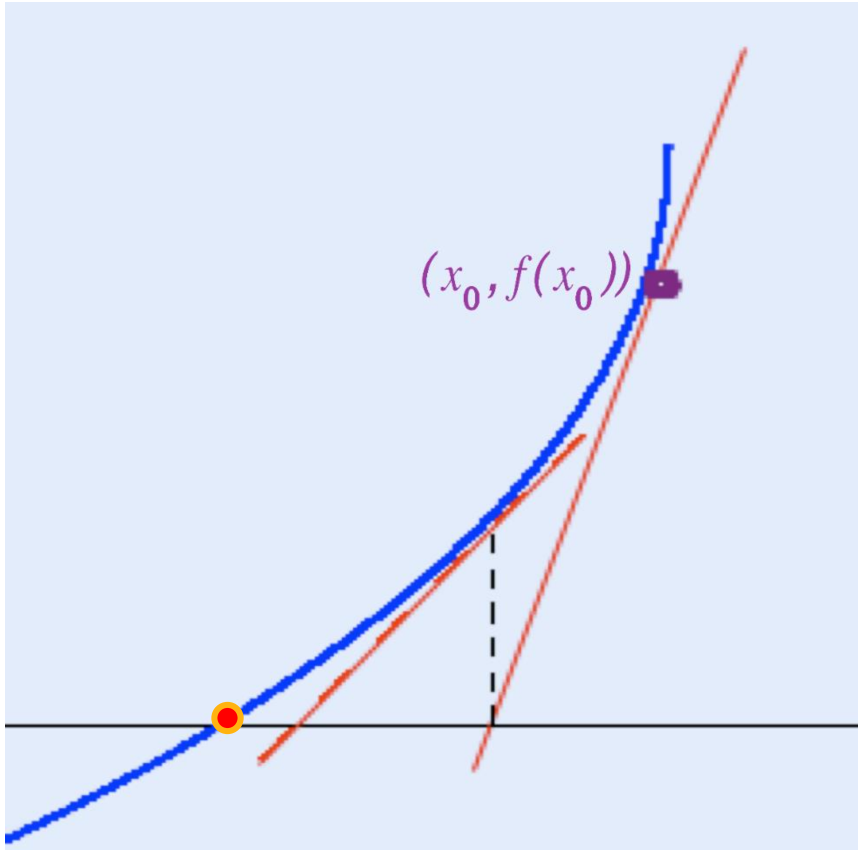
Suppose we are at the purple dotted point

We want to get to the red dotted point

What should we do?

Consider the successive X intercepts

Right !, successive x-intercepts of tangent lines get closer and closer to roots.



Newton's Method

Successive x-intercepts of tangent lines get closer and closer to roots.

Take a point $P_0 = (x_0, f(x_0))$

The x-intercept of the tangent at P_0 is $x_1 = x_0 - \frac{f(x_0)}{f\prime(x_0)}$ (check it out !)

The x-intercept of the tangent at $P_1 = (x_1, f(x_1))$ is $x_2 = x_1 - \frac{f(x_1)}{f\prime(x_1)}$ (refer above!)

Keep going: $x_0, x_1, x_2, x_3, \ldots.$

Some stopping criteria are:

- Limit on the number of iterations
- $|x_{n+1} - x_n| < \text{Limit}$

Example

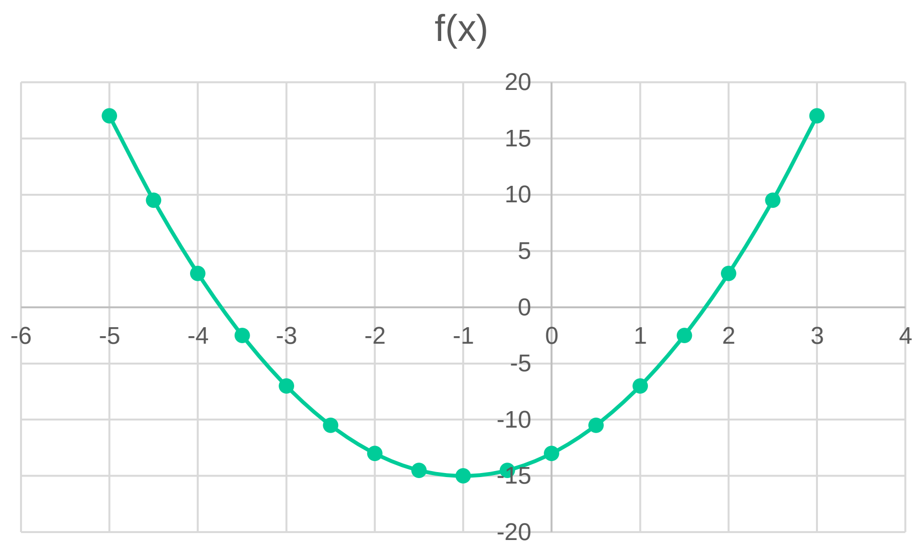
Find a root of $2x^2 + 4x - 13 = 0$

Take the initial approximation to be x = -2

$f'(x) = 4x + 4$

X intercept = $x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)} = x_i - \frac{2x_i^2 + 4x_i - 13}{4x_i + 4}$

x0 = -2, x1 = -5.25



#	0	1	2	3	4	5	6	7
x	-2	-5.25	-4.00735	-3.75062	-3.73864	-3.73861	-3.73861	-3.73861

Remarks

1. The method is not foolproof. It depends a lot on the initial guess.
2. The method can be extremely efficient, if the first guess is a good one.
3. The method is ideal for an Excel spreadsheet
4. The function may have no roots, the method will fail
5. x_n may not exist
6. The iterations may cycle

Consider $-x^3 + 4x^2 - 2x + 2$ with initial estimate $x_0 = 0$

Optimizing using Newton's Method

Minimize f(x)

The iterative method is:

$$x_{n+1} = x_n - \frac{f'(x_n)}{f''(x_n)}$$

Newton's method for finding minimum normally has quadratic convergence rate, but must be started close enough to solution to converge

Example

Min f(x) $= 0.5 - xe^{-x^2}$

f'(x) = $(2x^2 - 1)e^{-x^2}$

f''(x) = $2x(3 - 2x^2)e^{-x^2}$

And

$$x_{n+1} = x_n - \frac{f'(x_n)}{f''(x_n)} = x_n - \frac{(2x^2-1)e^{-x^2}}{2x(3-2x^2)e^{-x^2}}$$

#	x_n	f(x)	f'(x)	f''(x)	x_{n+1}
0	1.0000	0.1321	0.3679	0.7358	0.5000
1	0.5000	0.1106	-0.3894	1.9470	0.7000
2	0.7000	0.0712	-0.0123	1.7325	0.7071
3	0.7071	0.0711	-0.0001	1.7156	0.7071
4	0.7071	0.0711	0.0000	1.7155	0.7071

Steepest Descent Method

We want to minimize f(x)
Plan: Repeatedly compute the minima of a function g(t) of a single variable t, as follows:

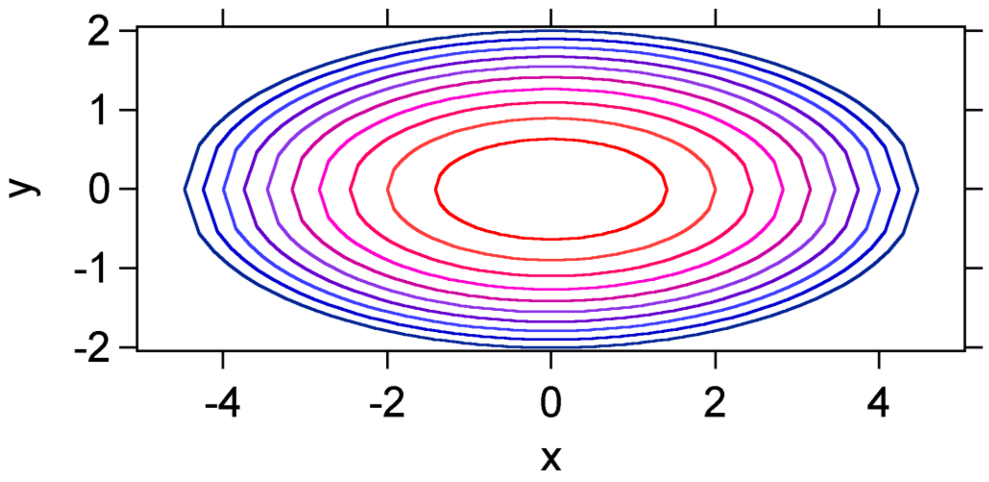
- Suppose f has a minimum at x.
- We start at a point x_0.
- We look for the minimum of f along the straight line in the direction $-\nabla f(x_0)$
  The direction of steepest decrease of f at x_0.
- x_1 is this point

Steps

Compute $\nabla f(x)$, $z(t) = x - t\nabla f(x)$ & g(t) = f(z(t))
Start at the initial guess x_0.
Find minima of g(t)
$x_1 = z(t) = x_0 - t\nabla f(x_0)$
Iterate until convergence

Steepest Descent Method – Intuition 1 / 2

$f(x, y) = x^2 + 5y^2$



$\nabla f(x) = [2x\ 10y]^T$

$z(t) = x - t\nabla f(x) = [x\ y]^T - t[2x\ 10y]^T$

$z(t) = [x - 2tx\ \ y - 10ty]^T$

$g(t) = f(z(t) = (x - 2tx)^2 + 5(y - 10ty)^2$

$= x^2(1 - 2t)^2 + 5y^2(1 - 10t)^2$

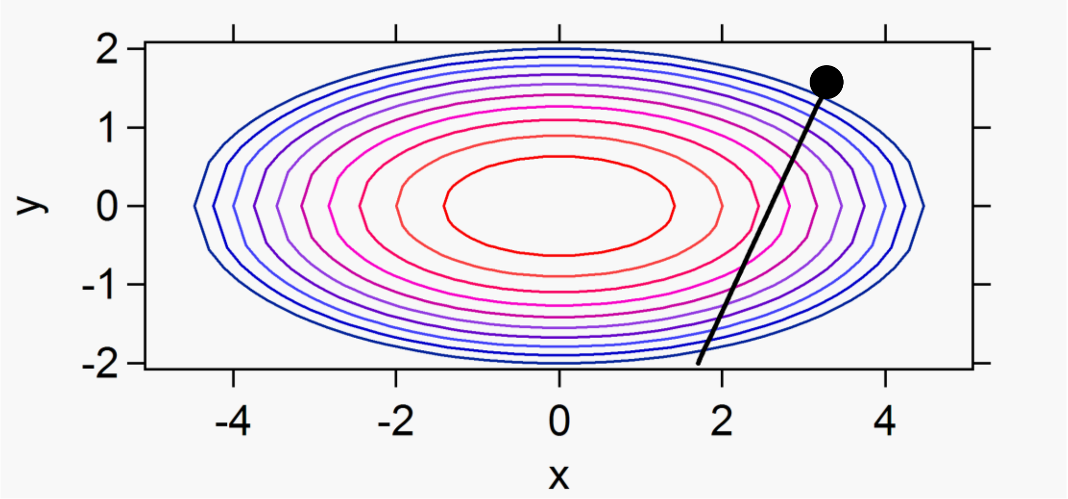
x_0: x = = 3.22 & y = 1.39

$\nabla f(x) = [2x\ 10y]^T = [6.44\ 13.9]^T$

The black dot is x_0

The black line is the direction of $\nabla f(x_0)$

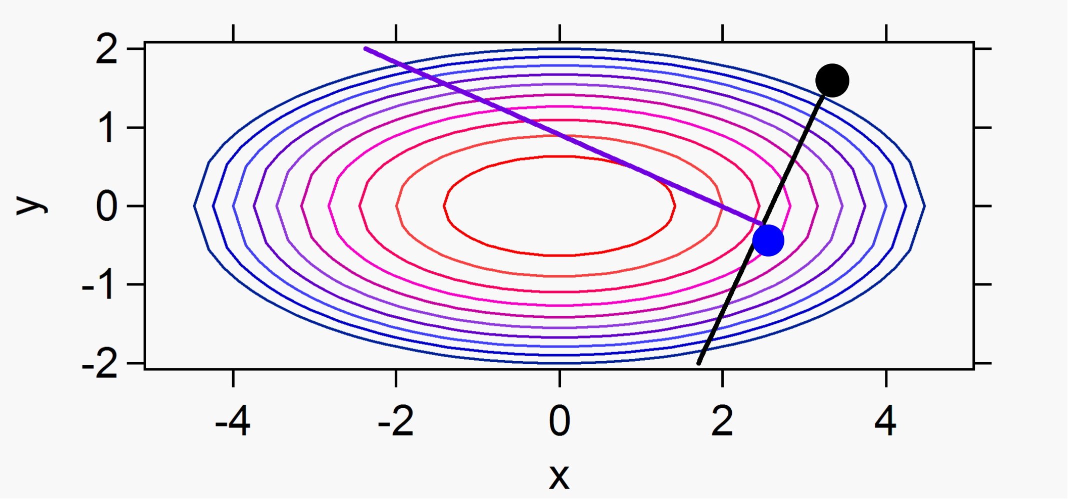
Contours: Red (f = 2) to Blue (f = 20)

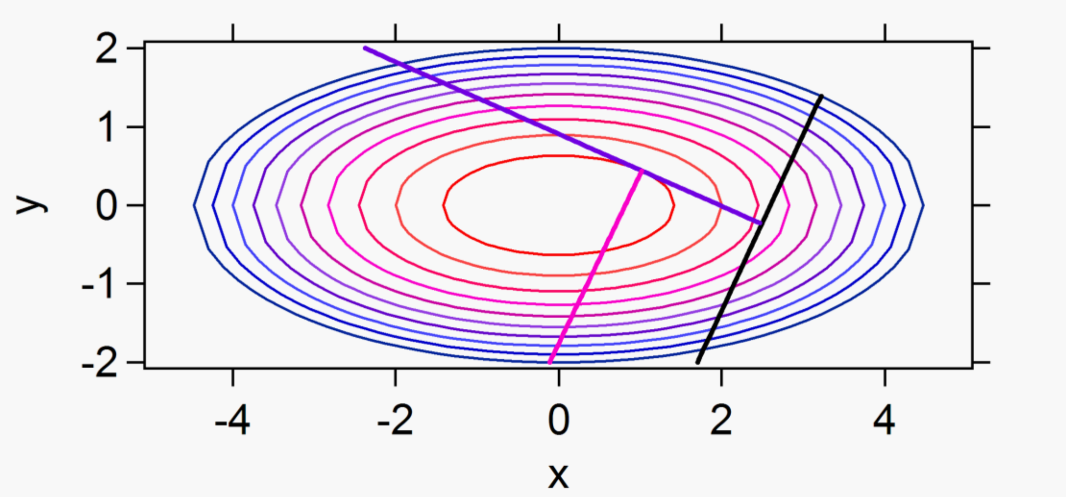


Steepest Descent Method – Intuition 2 / 2

How far should we go?

- The minimum along the line
- The current gradient is parallel to a contour





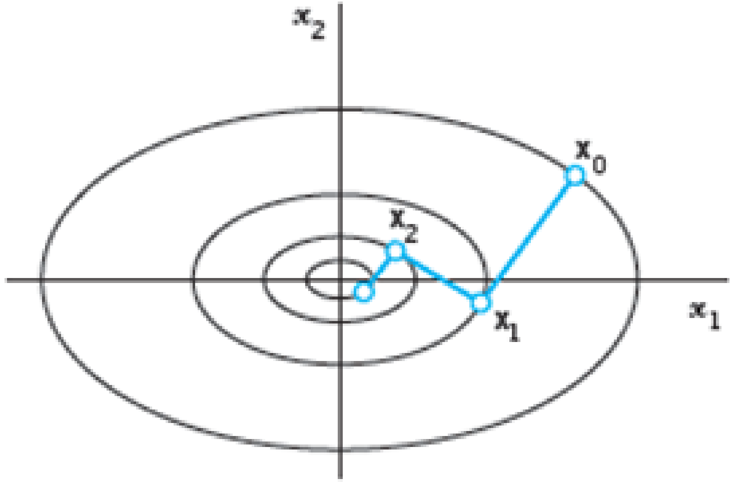
Example

Min $f(x, y) = x^2 + 3y^2$

$$\nabla f(x,y) = \begin{bmatrix} 2x \\ 6y \end{bmatrix}$$

$$z(t) = \begin{bmatrix} x \\ y \end{bmatrix} - t\nabla f(x,y)$$

$$= \begin{bmatrix} x \\ y \end{bmatrix} - t\begin{bmatrix} 2x \\ 6y \end{bmatrix} = \begin{bmatrix} (1-2t)x \\ (1-6t)y \end{bmatrix}$$



$g(t) = f(z(t)) = x^2(1-2t)^2 + 3y^2(1-6t)^2$

$g'(t) = -4x^2(1-2t) - 36y^2(1-6t)$

$g'(t) = 0 \Rightarrow t = \frac{4x^2+36y^2}{8x^2+216y^2}$

n	$\mathbf{x}$		t	$1-2t$	$1-6t$
0	6.000	3.000	0.210	0.581	−0.258
1	3.484	−0.774	0.310	0.381	−0.857
2	1.327	0.664	0.210	0.581	−0.258
3	0.771	−0.171	0.310	0.381	−0.857
4	0.294	0.147	0.210	0.581	−0.258
5	0.170	−0.038	0.310	0.381	−0.857
6	0.065	0.032			

Lagrange Multipliers

The constrained optima problem can be stated as:
Find the extreme value of z = f(x, y) subject to g(x, y) = 0

So Lagrange formed the augmented function. z = f(x, y) + λ g(x, y)

The first order conditions for optimization are:

$$\frac{\partial z}{\partial x} = 0, \frac{\partial z}{\partial y} = 0, \frac{\partial z}{\partial \lambda} = 0$$

- To be solved simultaneously
- These will give the critical points
- We will use intuitive methods to identify whether these critical points are maxima or minima

Example

Consider a firm producing commodity z with the following production function:

$z = 5xy$

Without any constraints, the firm can produce an unlimited quantity

But suppose the firm has a budget constraint:

Let the cost of the inputs x and y be: X = \$2 / unit and Y = \$1 / unit

For simplicity, assume that the maximum amount the firm can spend is \$100.

So we have the following constraint: $2x + y = 100$

Problem Formulations & Solution

Max 5xy subject to 2x + y = 100

The augmented function: z = 5xy + λ (2x + y -100)

The first order derivatives are:

$$\frac{\delta z}{\delta x} = 5y + \lambda 2; \frac{\delta z}{\delta y} = 5x + \lambda; \frac{\delta z}{\delta \lambda} = 2x + y - 100$$

Setting each derivative to 0 gives us: Three unknowns and 3 equation

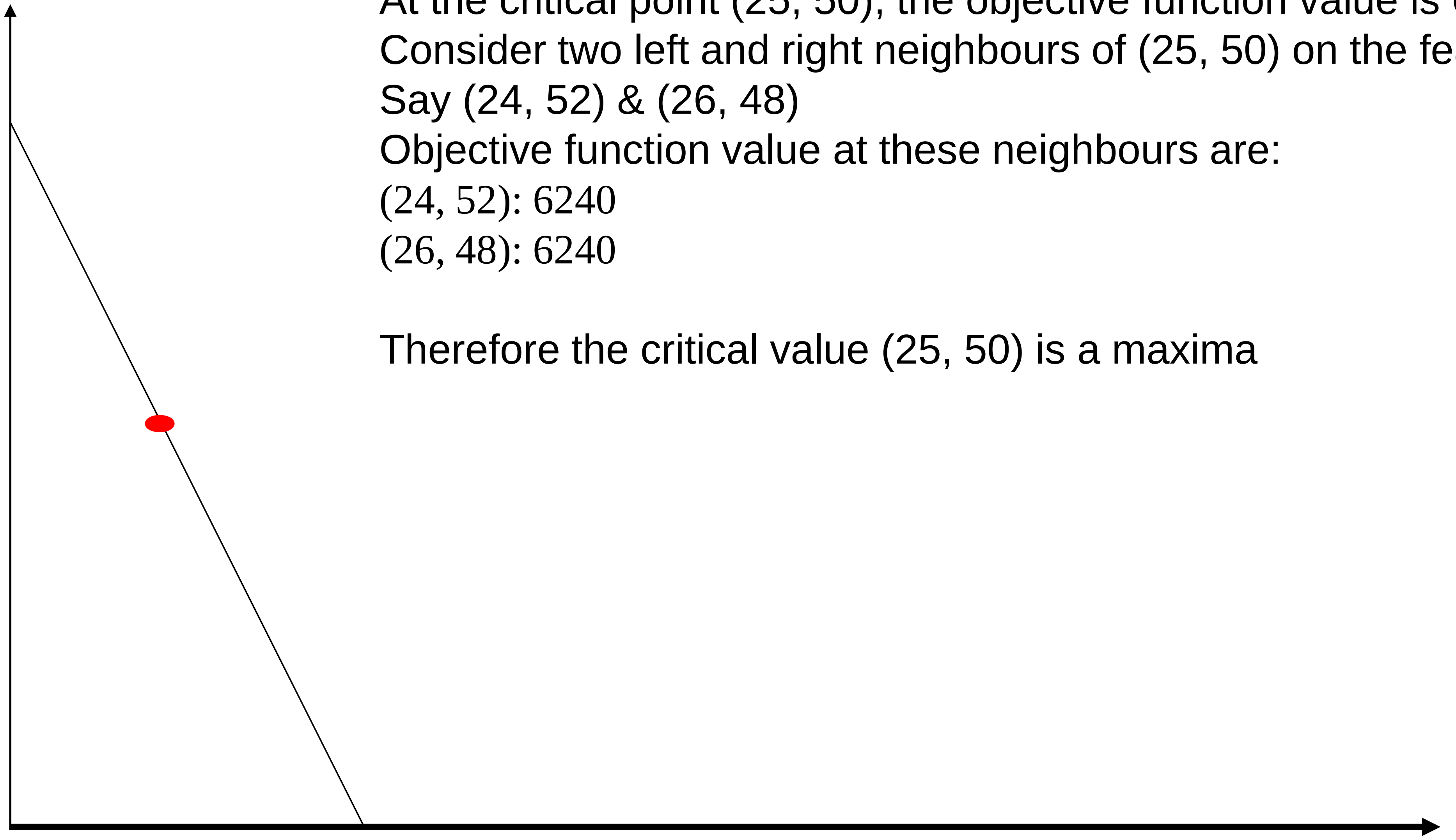
y = -2λ/5, x = -λ/5 & 2x + y -100 = -2λ/5 - 2λ/5 -100 = 0 → λ = 500 / 4 = 125

That is: x = 25 and y = 50 & 5xy = 6250

Is x = 25 and y = 50 a maxima?

Analysis of the Problem

Max 5xy subject to 2x + y = 100



At the critical point (25, 50), the objective function value is 6250
Consider two left and right neighbours of (25, 50) on the feasible space
Say (24, 52) & (26, 48)
Objective function value at these neighbours are:
(24, 52): 6240
(26, 48): 6240

Therefore the critical value (25, 50) is a maxima

Analytical Approach

There is an analytical way to show whether we have a maxima or a minima
But this is out of scope for this course