Principles and Practices of Data Science

Lecture 5

Melvin Ayala

Lecture 5: Vector Calculus and Optimization

Sections

1. Calculus in Computer Science

Calculus

- mathematical study of continuous change (like geometry deals with shapes and algebra with arithmetic operations)
- two major branches:
  - differential calculus (concerned with the calculation of instantaneous rates of change)
  - integral calculus (concerned with the summation of infinitely many small factors to determine some whole)
- relationship with Computer Science:
  - some topics are relevant, but not all.

Calculus 1	Calculus 2
Functions. New functions from old Trigonometric functions Exponential function, inverse functions, logarithms Derivative: motivation. Informal definition of limit Limit laws. Squeeze theorem Continuity, asymptotes Definition of derivative. Derivative as a function Derivative of polynomials. Product and quotient rules Derivatives of trig functions Chain rule, implicit differentiation Derivative of the logarithm Related rates, linear approximation Maximization. Mean value theorem Second derivative, convexity, second derivative test L'Hospital's rule L'Hospital's rule, more graph sketching Optimization problems Newton's method Antiderivatives Definite integral: definition The "area so far" function The fundamental theorem of calculus. Evaluating definite integrals via the "net change theorem" Substitution rule Areas between curves, average values	Review of integration, areas between curves Volumes Integration by parts Trigonometric integrals Trigonometric substitution Integration of rational functions Improper integrals Arc length Area of a surface of revolution Differential equations Separable equation Linear equations Parametrized curves Polar coordinates, area length in polar coordinates Sequences Series and the integral test Comparison tests Alternating Series, absolute convergence, ratio and root tests Strategy for testing series Power series, representations of functions as power series Taylor and Maclaurin series Applications of Taylor polynomials
Calculus 3	**Calculus 4**
Brief overview, coordinate systems Vectors Dot product Cross product Equations of lines and planes Parametric curves, conic sections Cylinders and quadric surfaces Vector functions and space curves Integrals of vector functions Arc length and curvature Motion: velocity and acceleration Functions of several variables, continuity Partial derivatives Tangent planes and linear approximations Chain rule Directional derivative, gradient Maximum and minimum Lagrange multipliers Complex numbers	Double integrals over general regions Polar coordinates, applications of double integrals More applications, triple integrals Cylindrical coordinates, spherical coordinates Spherical coordinates, change of variables Change of variables Vector fields Line integrals Fundamental theorem for line integrals Green's theorem Curl and divergence Parametric surfaces, surface area Surface integrals Stokes' Theorem Divergence Theorem Complex functions Cauchy-Riemann equations Contour integrals and Cauchy's Theorem

2. Partial Derivatives

Partial Derivative

- of a function of several variables
- derivative with respect to one of those variables
- other variables held constant
- partial derivatives are used in vector calculus and differential geometry
- change of rate of the function in one of the directions (dimensions)

Notations

- different notations: $\frac{\partial}{\partial x} f$, $\frac{\partial f}{\partial x}$

- we explicitly reflect when a function has different variables as $f(x, y, ...)$
 and we want to compute its partial derivatives with:

$$f(x, y, ...) = \frac{\partial}{\partial x} f(x, y, ...)$$

Partial Derivatives as a Limit

partial derivative also defined as a limit.

partial derivative of *f* at the point a = (a_1, a_2, …, a_n) with respect to the i-th variable x_i is defined as:

$$\frac{\partial}{\partial x_i} f(a) = \lim_{\mathrm{h}\to 0} \frac{f(a_1, a_2, \ldots, a_i + h, a_{i+1}, \ldots, a_n) - f(a_1, a_2, \ldots, a_i, a_{i+1}, \ldots, a_n)}{h}$$

$$\frac{\partial}{\partial x_i} f(a) = \lim_{\mathrm{h}\to 0} \frac{f(a + he_i) - f(a)}{h}$$

where e_i is a unit vector pointing in the direction of x_i.

Further Notations

it is common to specify explicitly which variables are being held constant.

the partial derivative of f with respect to x, holding y and z constant, is often expressed as:

$$\left(\frac{\partial f}{\partial x}\right)_{y,z}$$

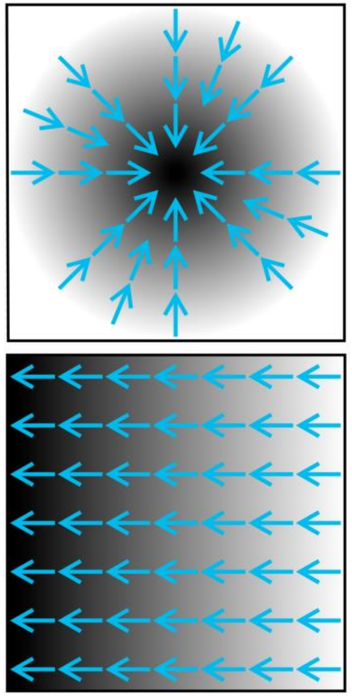
notation to evaluate partial derivative of a function at a specific point, e.g. (x, y, z) = (2, 4, 1):

$$\left(\frac{\partial f(x,y,z)}{\partial x}\right)\Bigg|_{(x,y,z)\,=\,(2,4,1)}$$

3. Gradients

Definition

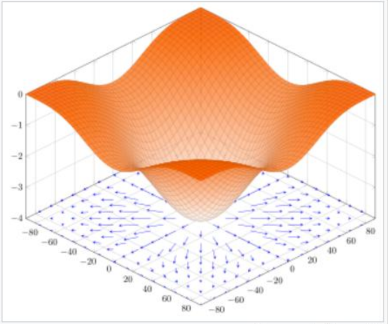
- **gradient** is a fancy word for derivative
- rate of change of a function.
- it's a vector (a direction to move) that:
    - points in the direction of greatest increase of a function
    - is zero at a local maximum or local minimum (because there is no single direction of increase)
- typically used for functions with several inputs and a single output (a scalar field)
- using "gradient" for single-variable functions is unnecessarily confusing



Source: Wikipedia

Properties

- the gradient of a multi-variable function has a component for each direction.
- a function that takes 3 variables will have a gradient with 3 components:
    - F(x) has one variable and a single derivative: dF/dx
    - F(x, y, z) has three variables and three partial derivatives: ∂F/∂x, ∂F/∂y, ∂F/∂z
- the gradient points in the direction of greatest increase (in multivariable functions, it can specify any direction).
- For a one variable function, there are no other components at all, so the gradient reduces to the derivative.



The gradient of the function $f(x,y) = -(\cos^2 x + \cos^2 y)^2$ depicted as a projected vector field on the bottom plane.

Source: Wikipedia

Notation

- It can be represented as vector of derivatives or vector of values (when evaluated at a point)
- given by the vector whose components are the partial derivatives of f at p.
- gradient of a multivariable function $f(x_1, x_2, \ldots, x_n)$ at point p is given by:

$$\nabla f(p) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(p) \\ \vdots \\ \frac{\partial f}{\partial x_n}(p) \end{bmatrix}$$

Other representations are common, e.g.:

$$Gradient\ of\ F(x,y,z) = \nabla F(x,y,z) = \left(\frac{\partial F}{\partial x}, \frac{\partial F}{\partial y}, \frac{\partial F}{\partial z}\right)$$

The gradient ∇f and the derivative df are expressed as a column and row vector:

$$\nabla f(p) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(p) \\ \vdots \\ \frac{\partial f}{\partial x_n}(p) \end{bmatrix}; \qquad df_p = \begin{bmatrix} \frac{\partial f}{\partial x_1}(p) & \cdots & \frac{\partial f}{\partial x_n}(p) \end{bmatrix}$$

Example

- Let's calculate the gradient of the following 3-dimensional function:

$$F(x, y, z) = x + y^2 + z^3$$

We get the following partial derivatives:

$$\frac{\partial F}{\partial x} = \frac{\partial}{\partial x}(x + y^2 + z^3) = 1$$

$$\frac{\partial F}{\partial y} = \frac{\partial}{\partial y}(x + y^2 + z^3) = 2y$$

$$\frac{\partial F}{\partial z} = \frac{\partial}{\partial z}(x + y^2 + z^3) = 3z^2$$

- therefore:

$$\nabla F(x, y, z) = \left(\frac{\partial F}{\partial x}, \frac{\partial F}{\partial y}, \frac{\partial F}{\partial z}\right) = (1, 2y, 3z^2)$$

- Let's find the direction of fastest increase at point (3, 4, 5)
- we must evaluate the gradient at that point

$$\nabla F(x, y, z)\Big|_{(3,4,5)} = (1, 2(4), 3(5)^2) = (1, 8, 75)$$

- therefore, vector (1, 8, 75) would be the direction we'd move in to increase the value of our function.

Notation in a Cartesian System

In the three-dimensional Cartesian coordinate system with a Euclidean metric, the gradient, if it exists, can be represented as:

$$\nabla f = \frac{\partial f}{\partial x}\,\boldsymbol{i} + \frac{\partial f}{\partial y}\,\boldsymbol{y} + \frac{\partial f}{\partial z}\,\boldsymbol{z}$$

where i, j, k are the standard unit vectors in the directions of the x, y and z coordinates, respectively. For example, the gradient of the function.

4. Jacobian and Hessian Matrices

4.1. Jacobian Matrix

Definition

- applied to vector-valued functions whose range is a set of multidimensional vectors.
- the Jacovian is the generalization of the gradient for vector-valued functions of several variables and differentiable Euclidean maps.

$$\boldsymbol{J} = \begin{bmatrix} \frac{\partial \boldsymbol{f}}{\partial x_1} & \frac{\partial \boldsymbol{f}}{\partial x_2} & \cdots & \frac{\partial \boldsymbol{f}}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} \\ \frac{\partial f_2}{\partial x_1} \\ \vdots \\ \frac{\partial f_m}{\partial x_1} \end{bmatrix} = \begin{bmatrix} \nabla^T f_1 \\ \nabla^T f_2 \\ \vdots \\ \nabla^T f_m \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

where $\nabla^T f_i$ is the transpose (row vector) of the gradient of the i-th component

The Jacobian matrix can be denoted in various ways: $D\boldsymbol{F}, \quad \boldsymbol{J_f}, \quad \nabla f, \quad and \quad \frac{\partial(f_1,\ldots,f_m)}{\partial(x_1,\ldots,x_n)}$

Usage

- to analyze the stability of a system

4.2. An Example

Find the Jacobian of the function F: $R^3 \rightarrow R^4$ with 4 functions and 3 variables as follows:

$y_1 = x_1$

$y_2 = 5x_3$

$y_3 = 4x_2^2 - 2x_3$

$y_4 = x_3 \sin x_1$

$$J_F(x_1, x_2, x_3) = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \frac{\partial y_1}{\partial x_3} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \frac{\partial y_2}{\partial x_3} \\ \frac{\partial y_3}{\partial x_1} & \frac{\partial y_3}{\partial x_2} & \frac{\partial y_3}{\partial x_3} \\ \frac{\partial y_4}{\partial x_1} & \frac{\partial y_4}{\partial x_2} & \frac{\partial y_4}{\partial x_3} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 5 \\ 0 & 8x_2 & -2 \\ x_3 \cos x_1 & 0 & \sin x_1 \end{bmatrix}$$

4.3. Hessian Matrix

Gradient and Hessian of a function

When a function in the Euclidean space is smooth enough, we can talk about its **gradient** and **Hessian**.

Gradient: $$\nabla f(x) = \left[\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \ldots, \frac{\partial f(x)}{\partial x n_1}\right]^T$$

All **second order derivatives** of the multivariate function f can be summarized in the so-called **Hessian** matrix whose element at row i and column j is:

$$\left(\nabla^2 f(x_1, x_2, \ldots, x_n)\right) = \frac{\partial^2 f(x_1, x_2, \ldots, x_n)}{\partial x_i \partial x_j}$$

In other words, $\nabla^2 f$ is the following n x n matrix:

$$\nabla^2 f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f_1}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

5. Convex Optimization

5.1. Preliminaries

Definition

Convex optimization:

- subfield of mathematical optimization
- studies the problem of minimizing convex functions over convex sets

Applications

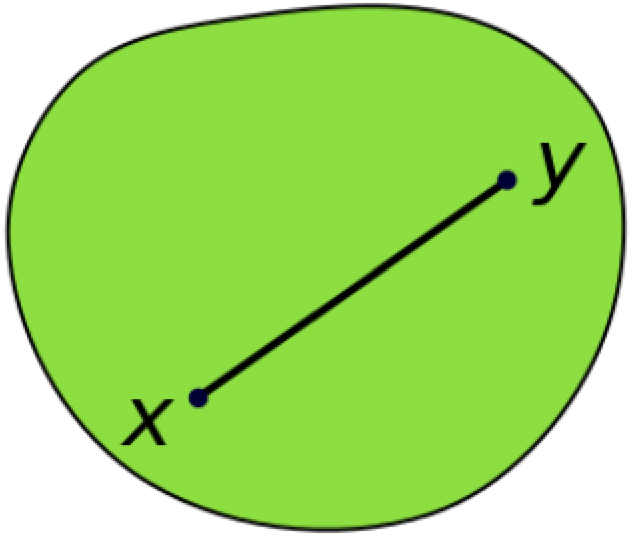
- control systems
- estimation
- signal processing,
- classification tasks
- communications and networks
- electronic circuit design
- finance
- statistics
- structural optimization

5.2. Convex Sets

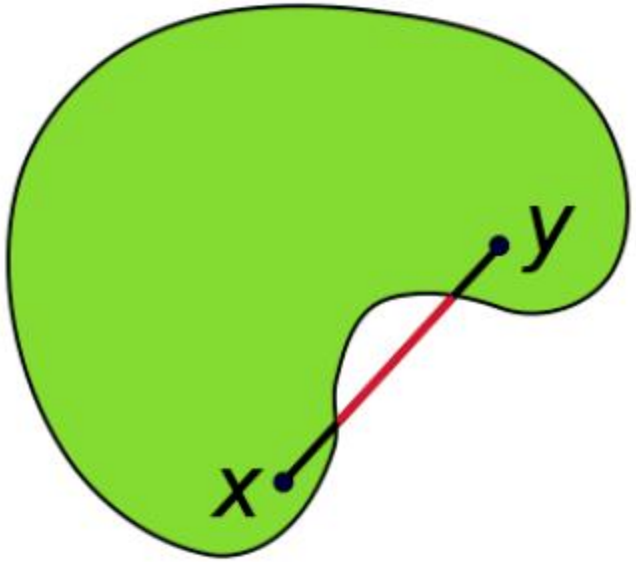
Definition

Convex sets:

- subset of Euclidean space
- a subset is convex if all points lying on a straight-line segment joining two points of the set are contained in the set.



convex subset



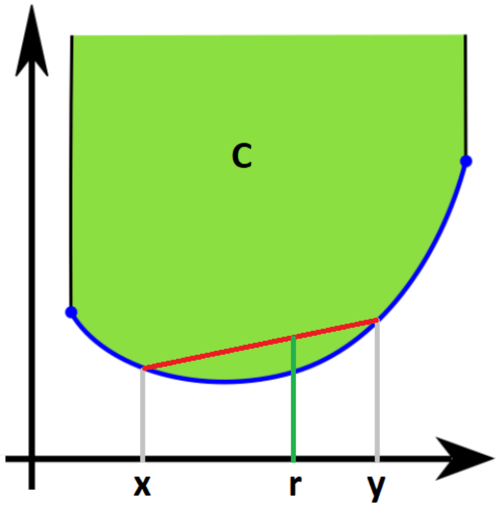
non-convex subset

5.3. Convex Functions

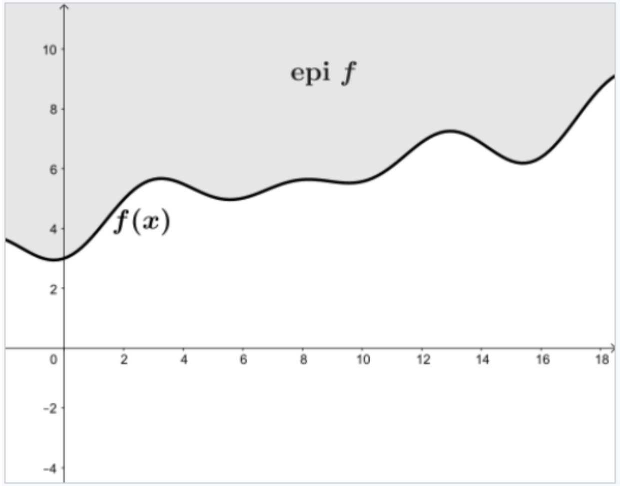
Definition:

A function is convex if and only if its epigraph, the region (in green) above its graph (in blue), is a convex set.

In mathematics, the epigraph or supergraph of a function is the set, denoted by epi f of all points in the Cartesian product lying on or above its graph.



$r = \lambda x + (1-\lambda)y$ belongs to C



Epigraph of a function

Convex Functions (cont.)

Definition:

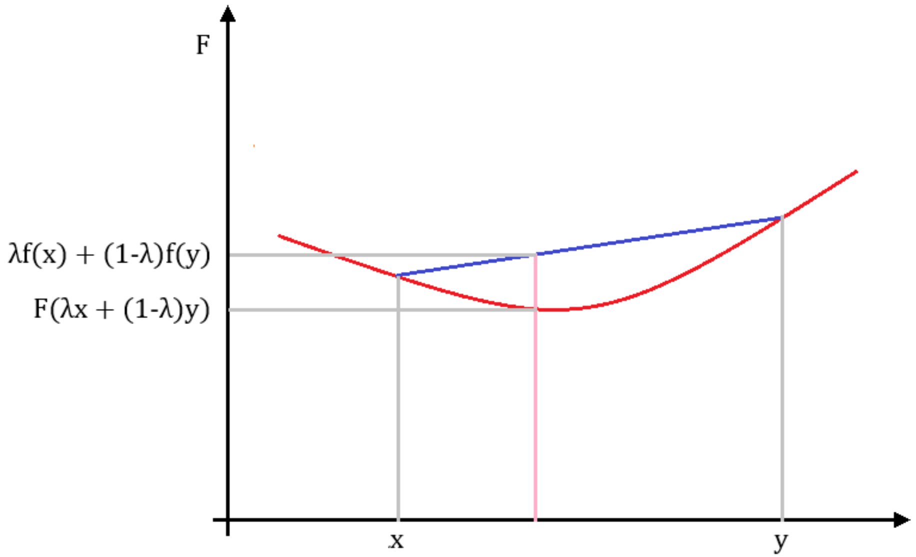
There are several ways to define a convex function.
A function is convex if $\forall x, y \in R$ and $\lambda \in [0, 1]$, we have:

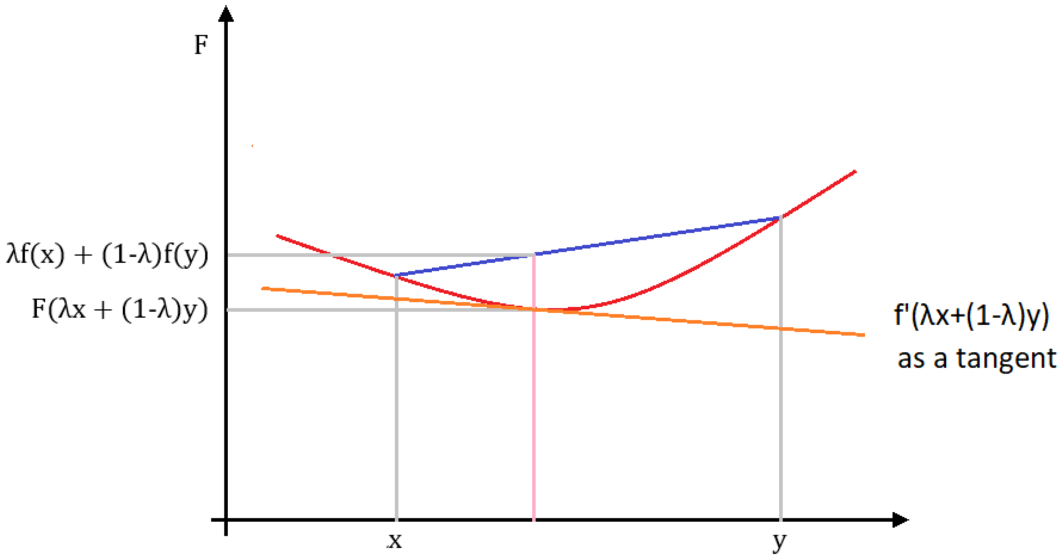
$$F(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$$

and therefore:

$$f(y) \geq f(x) + \langle \nabla f(x), y-x \rangle$$

(the tangent at f lies below f)





5.4. Algorithms for Convex Optimization

5.4.1. Overview

Optimization methods for functions:

- Gradient Descent
- Mirror Descent
- Multiplicative Weight Update Method
- Accelerated Gradient Descent
- Newton’s Method
- Interior Point Methods
- Cutting Plane and Ellipsoid Methods

Optimization methods for linear systems:

- Dantzig simplex algorithm (linear programming)

Examples:

- Maximize:

  $z = 8x_1 + 10x_2 + 7x_3$

  subject to the following constraints:

  (a) $x_1 + 3x_2 + 2x_3 \leq 10$

  (b) $-x_1 - 5x_2 - x_3 \geq 8$

  (c) $x_1, x_2, x_3 \geq 0$

- Maximize:

  $z = 8(x_1)^2 + 10x_2\, x_2 + 7(x_2)^2$

  subject to the following constraints:

  (a) $0.1x_1 + 0.1x_2 \leq 1$

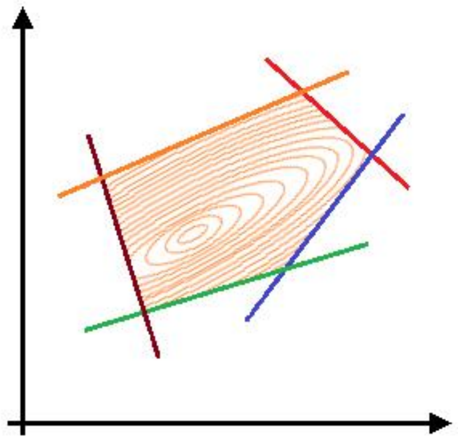
  (b) $(1/8)x_1 - (1/5)x_2 \leq 1$

  (c) $(1/3)x_1 - x_2 \leq 1$

  (d) $(1/4)x_1 + (1/6)\, x_2 - 1 \leq 0$

  (e) $-(1/5)x_1 + (1/5)\, x_2 - 1 \leq 0$

  (f) $x_1, x_2, x_3 \geq 0$



5.4.2. Newton's Method

Method and Purpose:

- named after Isaac Newton and Joseph Raphson
- also known as the Newton–Raphson
- is a root-finding algorithm
- produces successively better approximations to the roots of a real-valued function

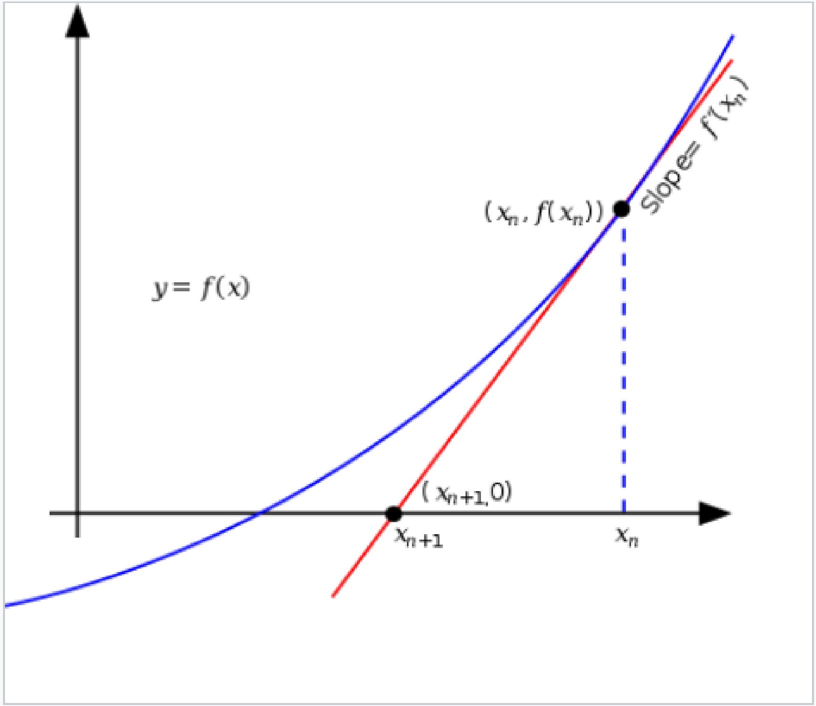
Idea of the Newton's Method:

- technique for solving equations of the form f(x)=0 by successive approximation.
- pick an initial guess x_0 such that $f(x_0)$ is reasonably close to 0. Then,

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$$

is a better approximation of the root than x_0.

(x1, 0): he intersection of the x-axis and the tangent of the graph of f at (x_0, $f(x_0)$)

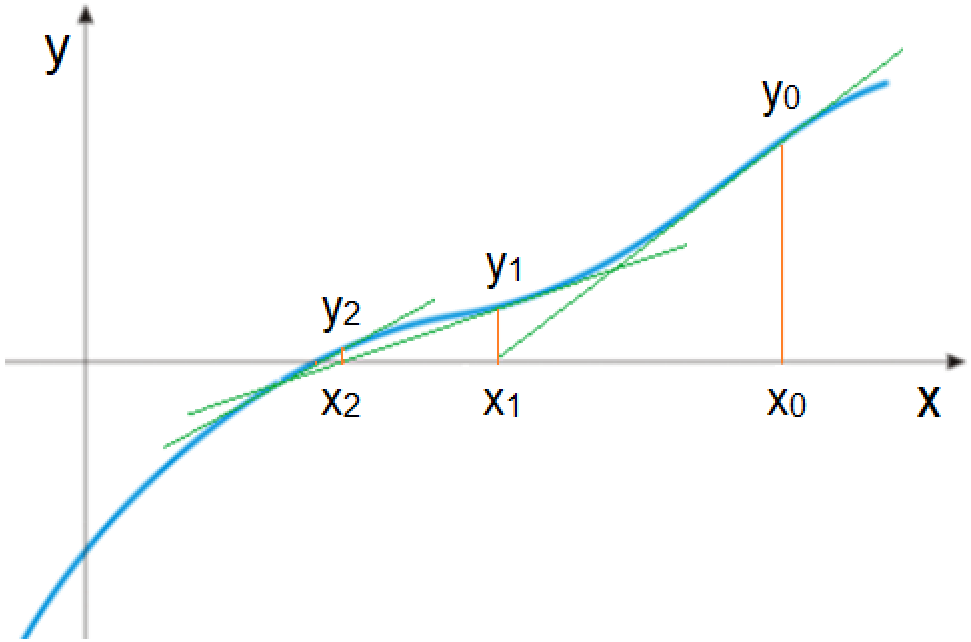


x_{n+1} is a better approximation than x_n for the root x of the function f (blue curve)

Newton's Method (Cont.)

The process is repeated as:

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}$$



Iteration usually improves the approximation.

Proof

Let's assume that the method converges, then:

- For a very small Δx, the secant gets closer to the tangent.
- If the root is in Δx, then the equation will yield a value very close to the root.

Slope at x_0 is computed as: $slope(x_0) = \frac{f(x_0)-0}{x_0 - x_1}$

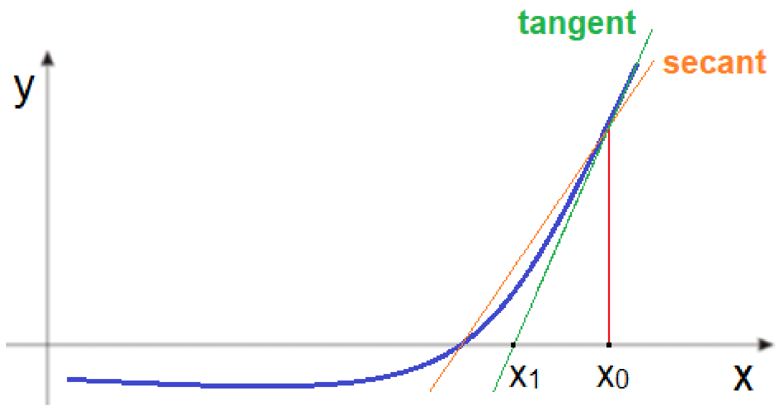
but slope(x_0) = f'(x_0), then: $f'(x_0) = \frac{f(x_0)-0}{x_0 - x_1}$

and therefore:

$$x_0 - x_1 = \frac{f(x_0) - 0}{f'(x_0)}$$

$$-x_1 = \frac{f(x_0)-0}{f'(x_0)} - x_0$$

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$$



Limitations of the Method

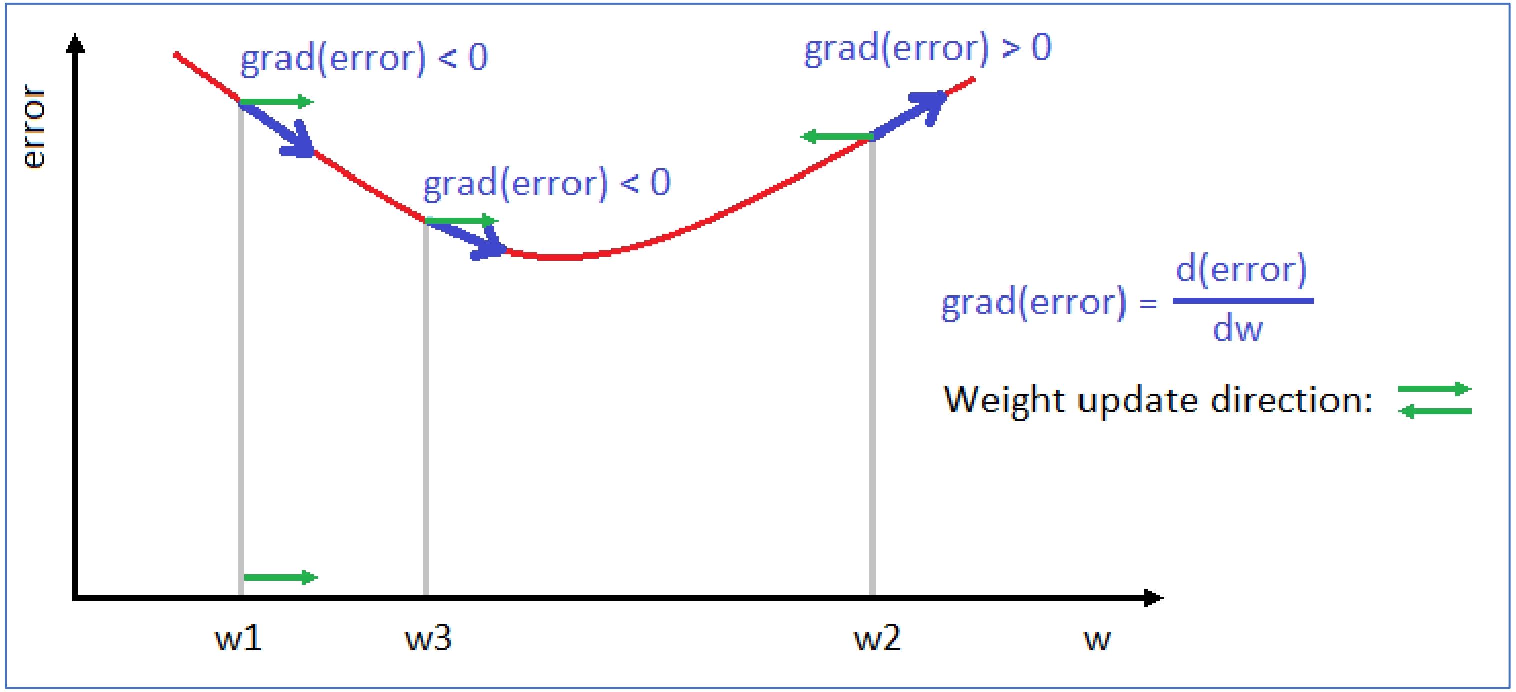
The Newton's method deserves some practical considerations:

- Difficulty in calculating the derivative of a function
- Failure of the method to converge to the root
- Overshoot
- Stationary point
- Poor initial estimate
- Mitigation of non-convergence

5.4.3. Gradient Descent Method

Gradient descent algorithm:

- iterative first-order optimization algorithm used to find a local minimum of a given function.
- commonly used in machine learning (ML) to minimize a cost/loss function (e. g. in a linear regression).



Rule:

When gradient is:

- negative, use a positive step.
- positive, use a negative step.

step = - gradient * LR

where LR: learning rate

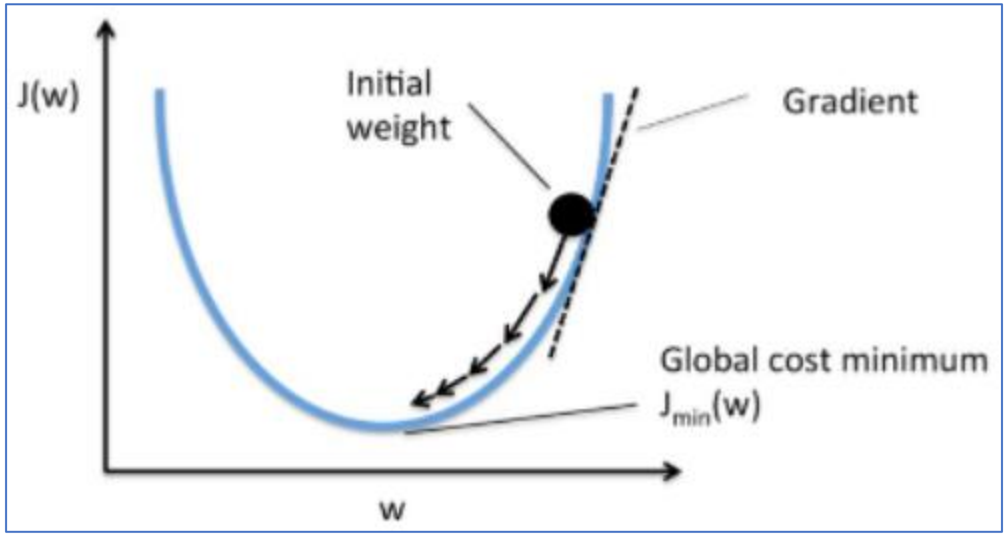
$w_{new} = w_{old} + step$

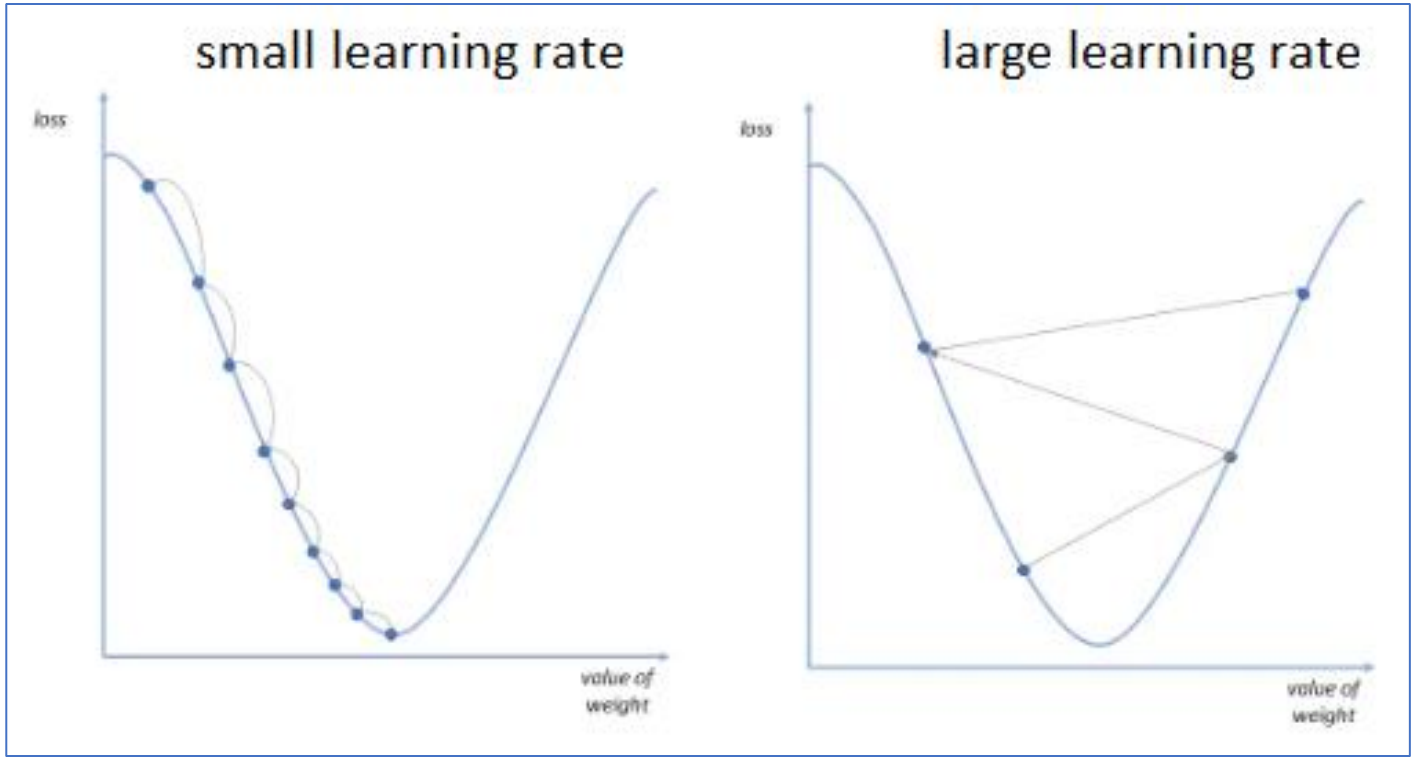
$w_{new} = w_{old} - grad(error) * LR$

Gradient Descent Method (cont.)

Learning Rate:

- decides how big the step will be
- Large LR: can create oscillations and miss the global minimum
- Small LR: can delay optimization but will find the optimum





5.4.4. Lagrange Multipliers Method

Method of Lagrange Multipliers:

- a strategy for finding the local maxima and minima of a function subject to equality constraints.
- named after the mathematician Joseph-Louis Lagrange.

The method can be summarized as follows:

- in order to find the maximum or minimum of a function f(x) subjected to the equality constraint g(x) = 0, form the Lagrangian function

$$\boldsymbol{L}(x, \lambda) = f(x) + \lambda g(x)$$

and find the stationary points of ***L*** considered as a function of x and the Lagrange multiplier λ

- all partial derivatives should be zero, including the partial derivative with respect to λ.
- the solution corresponding to the original constrained optimization is always a saddle point of the Lagrangian function.

Advantages:

- optimization solved without explicit parameterization
- widely used to solve challenging constrained optimization problems.
- generalized by the Karush–Kuhn–Tucker conditions, which can also take into account inequality constraints of the form $h(x) \leq c$ for a given constant c.

The Lagrange Multipliers Method With Equality Constraints

Suppose we have the following optimization problem:

Minimize f(x)

subject to: $g_1(x) = 0$, $g_2(x) = 0$, …, $g_n(x) = 0$

The method of Lagrange multipliers first constructs a function called the Lagrange function as given by the following expression:

$$L(x, \boldsymbol{\lambda}) = f(x) + \lambda_1 g_1(x) + \lambda_2 g_2(x) + \dots + \lambda_m g_m(x)$$

$$\text{with } \lambda = [\lambda_1, \lambda_2, \dots, \lambda_m]^T \quad \text{(Lagrange multipliers)}$$

To find the points of local minimum of f(x) subject to the equality constraints, we must solve:

$$\frac{\partial}{\partial x_i} L(x, \lambda) = 0 \qquad \text{for j=1, 2, …, n} \qquad \text{(number of variables)}$$

$$\frac{\partial}{\partial \lambda_j} L(x, \lambda) = 0 \qquad \text{for i=1, 2, …, m} \qquad \text{(number of constraints)}$$

We get a total of n+m equations to solve.

The Lagrange Multipliers Method With Equality Constraints: Example 1

Solve the following minimization problem:

Minimize: $f(x) = x^2 + y^2$

Subject to: $x + 2y - 1 = 0 \quad (= g(x, y))$

Solution:

The first step is to construct the Lagrange function:

$$L(x, y, \lambda) = f(x, y) + \lambda g(x, y)$$

$$L(x, y, \lambda) = x^2 + y^2 + \lambda(x + 2y - 1)$$

$\partial L/\partial x = 0$: $\quad 2x + \lambda = 0 \quad (1)$

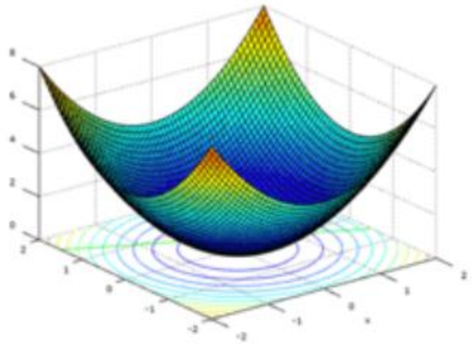
$\partial L/\partial y = 0$: $\quad 2y + 2\lambda = 0 \quad (2)$

$\partial L/\partial \lambda = 0$: $\quad x + 2y - 1 = 0 \quad (3)$

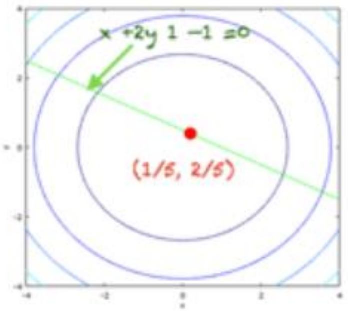
Using (1) and (2), we get: $\lambda = -2x = -y$

Plugging this in (3) gives: $\lambda = -(2/5)$

Plugging $\lambda = -(2/5)$ in (1) and (2) yields: $x = 1/5$ and $y = 2/5$



Graph of f(x, y) and the constraint



Contours of f(x, y) and the constraint

The Lagrange Multipliers Method With Equality Constraints: Example 2

Solve the following minimization problem:

Minimize: $f(x, y) = x^2 + 4y^2$

Subject to:

$x + y = 0 \quad (= g_1(x, y))$

$x^2 + y^2 = 0 \quad (= g_2(x, y))$

Solution:

The first step is to construct the Lagrange function:

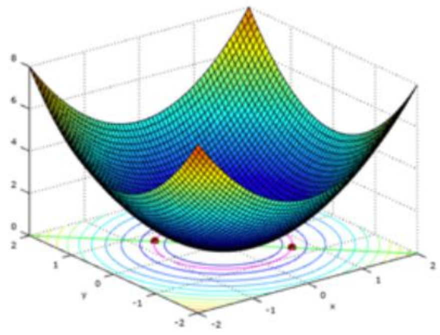
$$L(x, y, \lambda_1, \lambda_2) = f(x, y) + \lambda_1 g_1(x, y) + \lambda_1 g_2(x, y)$$

$$L(x, y, \lambda_1, \lambda_2) = x_2 + 4y_2 + \lambda_1(x + y) + \lambda_2(x_2 + y_2 - 1)$$

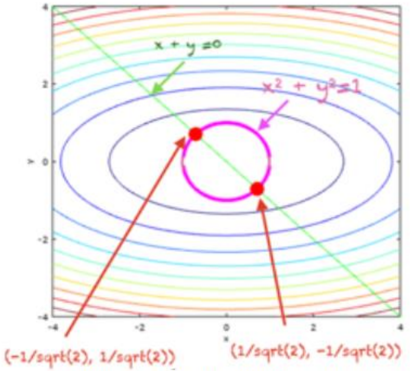
Equations to solve:

$\partial L/\partial x = 0$:	$2x + \lambda_1 + 2\lambda_2 x = 0$	(1)
$\partial L/\partial y = 0$:	$8y + \lambda_1 + 2\lambda_2 y = 0$	(2)
$\partial L/\partial \lambda_1 = 0$:	$x + y = 0$	(3)
$\partial L/\partial \lambda_2 = 0$:	$x_2 + y_2 - 1 = 0$	(4)

Solution: 2 points: (1/sqrt(2), -1/sqrt(2)) and (-1/sqrt(2), 1/sqrt(2))



Graph of f(x, y) and the constraint



Contours of f(x, y) and the constraint

Do We Need Equality Constraints?

Note that

$g(x) = 0 \quad \Leftrightarrow (g(x) \geq 0 \text{ and } g(x) \leq 0)$

Consider an equality-constrained problem:

minimize $f(x)$

subject to $g(x) = 0$

Can be rewritten as:

minimize $f(x)$

subject to $g(x) \geq 0$

$-g(x) \leq 0$

Equality constraints can be dropped for simplicity.

5.5. Duality Principle

- States that optimization problems may be viewed from either of two perspectives, the primal problem or the dual problem.

- If the primal is a minimization problem, then the dual is a maximization problem (and vice versa).

- Any feasible solution to the primal (minimization) problem is at least as large as any feasible solution to the dual (maximization) problem.

Duality Principle in Linear Programming

Consider an optimization problem of the form:

maximize $z = 120x + 100y$

subject to:

$2x + 2y \leq 8$

$5x + 3y \leq 15$

$x, y \geq 0$

The dual problem is:

minimize $z' = 8s + 15t$

subject to:

$2s + 5t \geq 120$

$2s + 3t \geq 100$

$s, t \geq 0$

Primal Problem

Dual Problem

Trick: Create new variables, swap coefficients, and change max to min and ≤ to ≥ (or vice versa).

Duality Principle in Linear Programming (Cont.)

Technique:

Write out the problem in short table form, flip it over, and write it back out in words.

2	2	8
5	3	15
120	100	

Dualize →

2	5	120
2	3	100
8	15	

Primal Problem

maximize $z = 120x + 100y$

subject to:

$2x + 2y \leq 8$

$5x + 3y \leq 15$

$x, y \geq 0$

Dual Problem

minimize $z' = 8s + 15t$

subject to:

$2s + 5t \geq 120$

$2s + 3t \geq 100$

$s, t \geq 0$

Duality Principle in Linear Programming (Cont.)

Another example:

5	6	7
8	9	10
3	4	

Dualize →

5	8	3
6	9	4
7	10	

Primal Problem

minimize $z = 3x + 4y$

subject to:

$5x + 6y \geq 7$

$8x + 9y \geq 10$

$x, y \geq 0$

↔

Dual Problem

maximize $z' = 7s + 10t$

subject to:

$5s + 8t \leq 3$

$6s + 9t \leq 4$

$s, t \geq 0$

Slack Variables in Linear Programming

Slack variables:

- additional variables that are introduced into the linear constraints of a linear program
- transform inequality constraints to equality constraints.

Consider the following linear programming problem:

$$\begin{aligned}
&\text{maximize } z = -x_1 - 2x_2\\
&\text{subject to:}\\
&x_1 - 2x_2 + x_3 \geq 4\\
&2x_1 + x_2 - x_3 \geq 6\\
&x_1, x_2, x_3 \geq 0
\end{aligned}$$

We change each constraint to an $\leq$ inequality and then introduce slack variables x_4 and x_5. The result is:

$$\begin{aligned}
&\text{maximize } z = -x_1 - 2x_2\\
&\text{subject to:}\\
&-x_1 + 2x_2 - x_3 + x_4 \qquad = -4\\
&-2x_1 - x_2 + x_3 \qquad + x_5 = -6\\
&x_1, x_2, x_3, x_4, x_5 \geq 0
\end{aligned}$$

End of Lecture