Gradient Descent

Move in the opposite direction of the gradient.

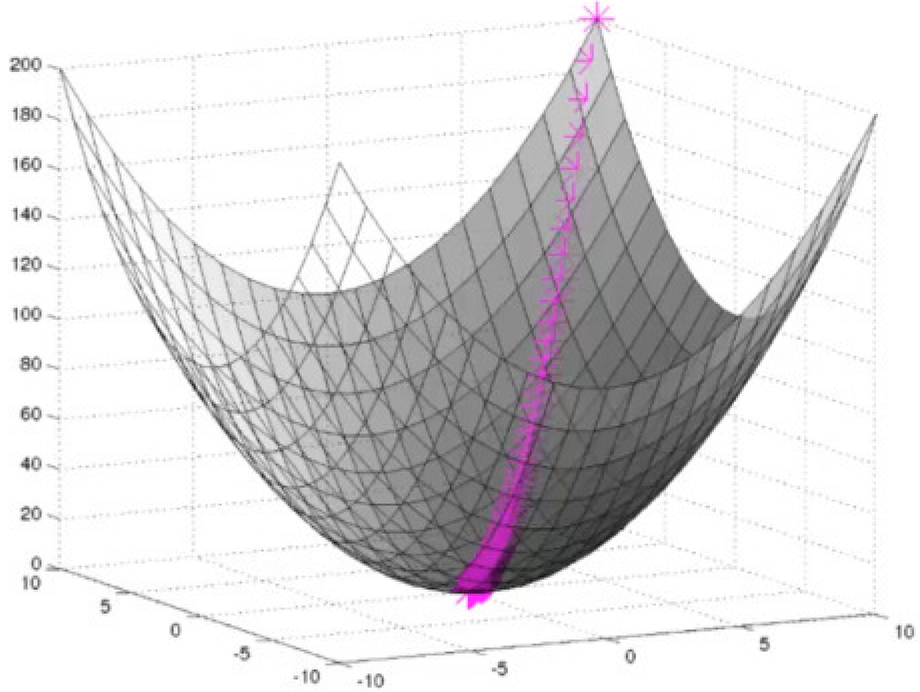
$$\Delta x = -\nabla f(x)$$

Algorithm 1 Gradient Descent

1: **Given** x_0 in dom(f)
2: **repeat**
3: $\quad \Delta x = -\nabla f(x)$
4: $\quad$ Update $x = x + t\Delta x$
5: **until** stopping criteria is satisfied

How do we choose t?
Is t constant?

Gradient Descent



Unconstrained Minimization

Consider a convex, twice differentiable function f then

$$\min f(x)$$

Assume, the minimum p^* is finite and is attained by f.
These algorithms produce a sequence of points x_i starting from a given point, such that

$$f(x_k) \to p^*$$

These algorithms require that the sublevel set at x_0 be closed.

Optimization Algorithms

There exists standard algorithms which can be used to solve optimization problems once in standard form. Some of them are

- Simplex Method for Linear Programs
- Interior Point methods

Optimization under no constraints

Various methods exist to solve this class of problems

- Gradient based methods
- Genetic Algorithms
- Simulated Annealing

Optimization Algorithms

There exists standard algorithms which can be used to solve optimization problems once in standard form. Some of them are

- Simplex Method for Linear Programs
- Interior Point methods

Example 2 (Contd..)

- Taking derivatives of L with respect to x_1, x_2 gives the following equations.

$$x_1 = \frac{\lambda_1 + \lambda_2}{1 + \lambda_1 + \lambda_2}$$

$$x_2 = \frac{\lambda_1 - \lambda_2}{1 + \lambda_1 + \lambda_2}$$

- Subsituting them in the Lagrangian we get

$$g(\lambda_1, \lambda_2) = \frac{\lambda_1 + \lambda_2 + (\lambda_1 - \lambda_2)^2}{1 + \lambda_1 + \lambda_2}$$

  We see it is symmetric and substitute $\lambda_1 = \lambda_2$ to get,

$$g(\lambda_1, \lambda_2) = \frac{2\lambda_1}{2\lambda_1 + 1}$$

- $g(\lambda_1, \lambda_2) \to 1$ as $\lambda_1 \to \infty$. $p^* = d^* = 1$. KKT not satisified, Slater's condition not satisfied.

Example 2 (Contd..)

- Let us now list the KKT conditions,

$$
\begin{aligned}
(x_1-1)^2+(x_2-1)^2 &\leq 1\\
(x_1-1)^2+(x_2+1)^2 &\leq 1\\
\lambda_1 &\geq 0\\
\lambda_2 &\geq 0\\
2x_1+2\lambda_1(x_1-1)+2\lambda_2(x_1-1) &= 0\\
2x_2+2\lambda_1(x_2-1)+2\lambda_2(x_2+1) &= 0\\
\lambda_1[(x_1-1)^2+(x_2-1)^2-1] &= 0\\
\lambda_2[(x_1-1)^2+(x_2+1)^2-1] &= 0
\end{aligned}
$$

- At $(1,0)$ the equations are not valid.

Example 2

$$\begin{array}{lrcl} \min & x_1^2 + x_2^2 & & \\ \text{s.t.} & (x_1-1)^2 + (x_2-1)^2 & \leq & 1 \\ & (x_1-1)^2 + (x_2+1)^2 & \leq & 1 \end{array}$$

where $x \in \mathbb{R}^2$.

- We can see analytically that the each of the constraints define circular regions with centers at $(1,1)$ and $(1,-1)$ of radius 1. There is only one point in common which is $(1,0)$. $p^* = 1$.
- Lagrangian,

$$L(\bar{x}, \bar{\lambda}) = x_1^2 + x_2^2 + \lambda_1((x_1-1)^2 + (x_2-1)^2 - 1) +$$

$$\lambda_2((x_1-1)^2 + (x_2+1)^2 - 1)$$

Example 1 - Least Squares

$$\min ||Ax - b||_2^2$$

- Convex function with no constraints.
- Can be solved analytically to give $x^* = (A^T A)^{-1} A^T b$
- Very commonly occurring structure, eg. Linear Regression.

KKT conditions

- If x, λ, ν satisfy strong duality then KKT conditions hold. They are necessary conditions for a solution to be optimal.
- For a problem where Slater's conditions are satified, KKT conditions become sufficient too.

Karush-Kuhn-Tucker Conditions

The following 4 conditions are known as KKT conditions (for the standard problems where f_i, h_i are differentiable.

- Stationarity

$$\nabla(f_0(x^*) + \sum_{i=0}^{m} \lambda_i^* f_i(x^*) + \sum_{i=0}^{p} \nu_i^* h_i(x^*)) = 0$$

- Primal feasibility

$$f_i(x^*) \leq 0, i = 1, 2, 3, \ldots m$$

$$h_i(x^*) = 0, i = 1, 2, 3, \ldots p$$

- Dual feasibility

$$\lambda_i^* \geq 0, i = 1, 2, 3, \ldots m$$

- Complementary slackness

$$\lambda_i^* f_i(x^*) = 0, \forall i = 1, 2, 3, \ldots m$$

Complementary slackness

Assume strong duality holds, x^* is primal optimal and (λ^*, ν^*) are dual optimal, then we can write the following

$$\begin{aligned} f_0(x^*) = g(\lambda^*, \nu^*) &= \inf_x \left(f_0(x) + \sum_{i=1}^{i=m} \lambda_i^* f_i(x) + \sum_{i=1}^{i=p} \nu_i^* h_i(x)\right) \\ &\leq f_0(x^*) + \sum_{i=1}^{i=m} \lambda_i^* f_i(x^*) + \sum_{i=1}^{i=p} \nu_i^* h_i(x^*) \\ &\leq f_0(x^*) \end{aligned}$$

From these equations, we can say the following,

- x^* minimizes $L(x, \lambda, \nu)$
- $\lambda_i^* f_i(x^*) = 0$ for all $i = 1, 2, \ldots m$

$$\lambda_i^* > 0 \Rightarrow f_i(x^*) = 0; f_i(x^*) < 0 \Rightarrow \lambda_i^* = 0$$

Slater's Conditions

Sufficiency condition for Strong Duality

In a convex optimization problem, slater's condition implies that if $\exists x \in \text{reint}\mathbb{D}$ such that $f_i(x) < 0$ and $h_i(x) = 0$ then strong duality holds.

In other words, slater condition states that, strong duality holds if there exists a point x in the interior of feasible region of the problem.

Strong and Weak Duality

If the duality gap is 0, then it is known as Strong Duality.
The primal problem can also be written as

$$p^* = \inf_x \sup_{\lambda \geq 0, \nu} L(x, \lambda, \nu)$$

The dual problem can be written as

$$d^* = \sup_{\lambda \geq 0, \nu} \inf_x L(x, \lambda, \nu)$$

If strong duality holds, we can see that the order of inf and sup don't matter. The optimal variables occur at a saddle point of the Lagrangian.

Lagrangian Dual problem

$$g(\lambda, \nu) \leq p^*$$

g forms a lower bound on the optimal value of the primal problem.

Dual problem

$$\begin{array}{lll} \max & g(\lambda, \nu) & \\ \text{s.t.} & \lambda_i \geq 0 & i = 1, 2, 3, \ldots m. \end{array}$$

The optimal value of this problem is attained at λ^*, ν^*.
We can see that the dual is concave irrespective of the form of the primal problem and can be solved. The optimal solution of the dual problem is denoted by d^*.

$$p^* - d^*$$

is known as the duality gap.

Lagrangian

Let us consider an alternative relaxed problem,

$$\begin{array}{llll} \min & f_0(x) + \sum \lambda_i f_i(x) + \sum \nu_i h_i(x) & & \\ \text{s.t.} & \lambda_i \geq 0 & i = 1, 2, 3, \ldots m & (5) \\ & x \in D & & \end{array}$$

$$L(x, \lambda, \nu) = (f_0(x) + \sum_{i=0}^{m} \lambda_i f_i(x) + \sum_{i=0}^{p} \nu_i h_i(x))$$

$$\begin{aligned} \inf_x (f_0(x) + \textstyle\sum_{i=0}^{m} \lambda_i f_i(x) + \sum_{i=0}^{p} \nu_i h_i(x)) & \leq L(x*, \lambda, \nu) \\ & \leq p^* \end{aligned}$$

$$g(\lambda, \nu) = \inf_x L(x, \lambda, \nu)$$

Duality

Every problem can be seen in two perspectives, the *primal form* and *dual form*.
Solving and understanding the dual helps us understand the behaviour of the primal form.
Consider the standard form,

$$\begin{array}{llll} \min & f_0(x) & & \\ \text{s.t.} & f_i(x) \leq 0 & & i = 1, 2, 3, \dots m, \\ & a_i^T x = b_i & & i = 1, 2, 3, \dots p. \end{array} \tag{4}$$

Let $\mathbb{D}$ denote the domain of the problem. This problem is called the primal problem, and it's optimal value is denoted by p^* obtained at x^*.

Convex Optimization

$$\begin{aligned} \min \quad & f_0(x) & & \\ \text{s.t.} \quad & f_i(x) \leq 0 & & i = 1, 2, 3, \ldots m, \\ & a_i^T x = b_i & & i = 1, 2, 3, \ldots p. \end{aligned} \tag{3}$$

- f_0 should be convex
- f_i should be convex
- Equality constraints should be affine.

Observation- The domain becomes convex always for this problem.
Why are convex optimization problems interesting?

Optimization Problem

Consider a general optimization problem,

$$\begin{array}{llll} \min & f_0(x) & & \\ \text{s.t.} & f_i(x) \leq 0 & & i = 1, 2, 3, \ldots m, \\ & h_i(x) = 0 & & i = 1, 2, 3, \ldots p. \end{array} \tag{2}$$

The optimal value p^* is given by

$\inf\{f_0(x) | f_i(x) \leq 0, i = 1, 2, \ldots, m \text{ and } h_i(x) = 0, i = 1, 2, 3, \ldots, p\}$

Why the *infimum* and not *min*?

Jensen's Inequality

For a convex function f, and $x_1, x_2, \ldots, x_n$ in it's domain, and let $\theta_1, \theta_2, \ldots, \theta_n$ such that $\theta_i \geq 0, \forall i$ and $\sum_{i=0}^{n} \theta_i = 1$.

$$f(\theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n) \leq \sum_{i=0}^{n} \theta_i f(x_i)$$

Value of the average is less than the average of the values.

Properties

- Convexity Preserving Operations
  - Non-negative Weighted Sum
    $\sum \alpha_i f_i$ is convex if $\alpha_i \geq 0$
  - Composition with Affine Function
    $f(Ax + b)$ is also convex if f is convex.
  - Pointwise Maximum and Supremum

$$f_1, f_2 \text{ convex} \implies \max\{f_1(x), f_2(x)\} \text{ convex}$$

  - Minimization
    If $f(x, y)$ is convex in (x, y) and C is a convex set, then
    $g(x) = inf_{y \in C} f(x, y)$ is convex
- Local minima is the global minima.

Sublevel sets

Sublevel sets

The (α-) sublevel set of f is

$$C(\alpha) \triangleq \{x \in \mathrm{dom}(f) | f(x) \leq \alpha\}$$

f convex $\implies$ sublevel sets are convex

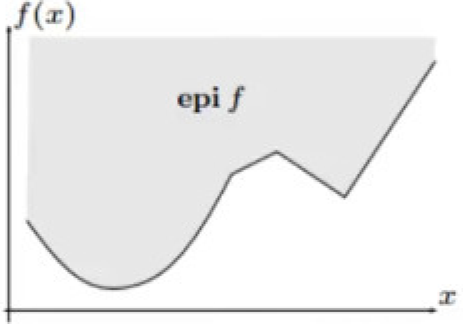
Converse is not true.

Epigraph

Epigraph

$$\text{epi } f = \{(x,t) | x \in \text{dom}(f), t \geq f(x)\}$$

f is a convex function $\Leftrightarrow$ epi f is convex set.



Second Order Condition

Let f be twice differentiable, i.e, $\nabla^2 f$ (its Hessian or second derivative) exists for each x in dom(f). f is convex if and only if:

- dom(f) is convex
- Its Hessian is positive semi-definite , i.e, $\nabla^2 f(x) \succeq 0, \forall x \in \text{dom}(f)$

We notice that in $\mathbb{R}$, the condition reduces to $f'' \geq 0$.

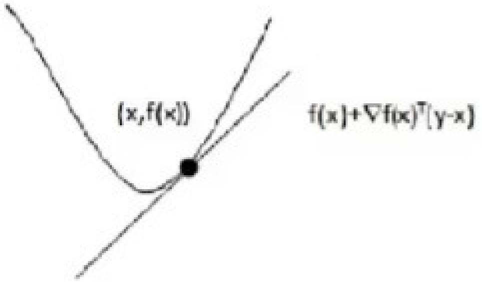
$$\mathbf{H} = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1\,\partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_1\,\partial x_n} \\ \dfrac{\partial^2 f}{\partial x_2\,\partial x_1} & \dfrac{\partial^2 f}{\partial x_2^2} & \cdots & \dfrac{\partial^2 f}{\partial x_2\,\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f}{\partial x_n\,\partial x_1} & \dfrac{\partial^2 f}{\partial x_n\,\partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_n^2} \end{bmatrix}.$$

Conditions for Convexity

First Order Condition

Let f be differentiable, i.e, ∇f exists for each x in $\mathrm{dom}(f)$. f is convex if and only if:

- $\mathrm{dom}(f)$ is convex.
-
$$f(y) \geq f(x) + \nabla f(x)^T (y - x), \forall x, y \in \mathrm{dom}(f)$$



Strictly Convex Functions

$$f(\theta x + (1-\theta)y) < \theta f(x) + (1-\theta) f(y)$$

, when $x \neq y$.

Concave Function
A function f is said to be concave if $-f$ is convex.

Strictly Concave Function
A function f is said to be strictly concave if $-f$ is strictly convex.

Convex Function

Definition

A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be convex if,

- Domain of f is a convex set
- $\forall x, y \in \mathrm{dom}(f)$, and $0 \leq \theta \leq 1$, we have

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y)$$

Example

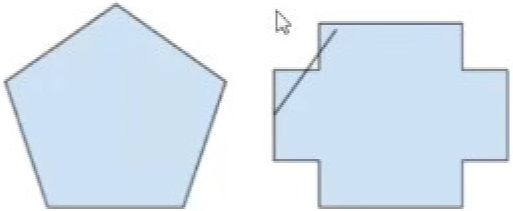


Figure: Convex Set(Left); Non-convex Set(Right)

Convex Set

Definition

A set C is convex if for all points $a, b \in C$ then the line segment through the points a, b lies in the set C, i.e., $c = \theta a + (1 - \theta)b$, $c \in C, \forall \theta \in [0, 1]$

Convex Combination

A point of the form $\theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k$ such that $\sum_{i=1}^{k} \theta_i = 1$ and $\theta_i \geq 0$ is known as the convex combination of the k points $x_1, x_2, x_3, \ldots x_k$.

Solving Optimization problems

- Optimizations are very tough problems to solve.
- Optimization problems are classified into various classes based on the properties of objectives and constraints.
- Some of these classes can be solved efficiently
  - Linear programs
  - Least Squares problems
  - Convex Optimization problems
- We will study Convex optimization problems, as we come across these problems very regularly.

Examples

- Data fitting:
  - **Variables**: Parameters of the model
  - **Constraints**: Parameter limits, prior information.
  - **Objective**: Measure of fit (Eg. Minimizing of error).
- Portfolio Optimization:
  - **Variables**: amounts invested in different assets
  - **Constraints**: budget, max./min. investment per asset, minimum return
  - **Objective**: overall risk or return variance

Optimal Solution

When do I know any $x \in \mathbb{R}^n$ is the solution for the problem?

- x satisfies all the constraints
- $f_0(x)$ is the minimum possible value in the feasible region.

Such a vector is generally represented by x^*, called as optimal solution.

Mathematical Optimization

Definition

Mathematical optimization is the selection of a best element (with regard to some criteria) from some set of available alternatives.

$$\begin{array}{lllll} \min & f_0(x) & & & \\ \text{s.t.} & f_i(x) & \leq & b_i & i = 1, 2, 3, \ldots m, \end{array} \tag{1}$$

where, $x \in \mathbb{R}^n$ known as the optimization variable
$f_0 : \mathbb{R}^n \to \mathbb{R}$ defines the criteria. Also known as the objective function
$f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, 2, 3, \ldots m$ are known as the contraints.