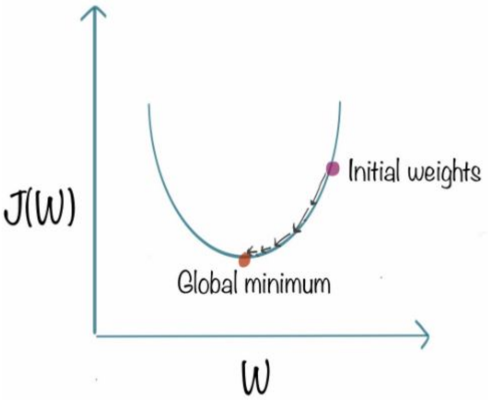
Chapter 7. Continuous optimization

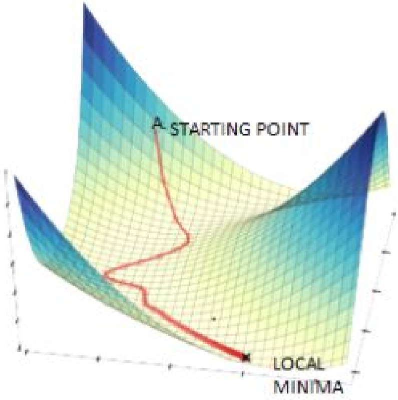
FPT UNIVERSITY

Chapter 7. Continuous optimization

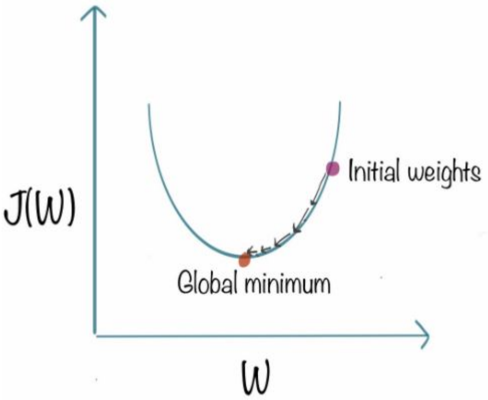
FPT Education
FPT UNIVERSITY

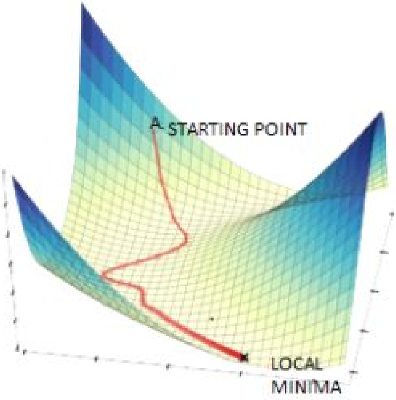
7.1 Optimization Using Gradient Descent





7.1 Optimization Using Gradient Descent





Consider the problem

$$\min_x f(x),$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable.

Gradient descent is a first-order optimization algorithm. To find a local minimum of a function using gradient descent, one takes steps proportional to the negative of the gradient of the function at the current point.

Consider the problem

$$\min_x f(x),$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable.

Gradient descent is a first-order optimization algorithm. To find a local minimum of a function using gradient descent, one takes steps proportional to the negative of the gradient of the function at the current point.

We can see that, if

$$x_1 = x_0 - \gamma\left((\nabla f)(x_0)\right)^T$$

for a small step size (learning rate) $\gamma > 0$ then $f(x_1) \leq f(x_0)$.

Consider the problem

$$\min_x f(x),$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable.

Gradient descent is a first-order optimization algorithm. To find a local minimum of a function using gradient descent, one takes steps proportional to the negative of the gradient of the function at the current point.

We can see that, if

$$x_1 = x_0 - \gamma \left((\nabla f)(x_0) \right)^T$$

for a small step size (learning rate) $\gamma > 0$ then $f(x_1) \leq f(x_0)$.

Iterate

$$x_{i+1} = x_i - \gamma_i \left((\nabla f)(x_i) \right)^T.$$

For suitable learning rate γ_i, the sequence $f(x_0) \geq f(x_1) \geq \cdots$ converges to a local minimum.

Consider the problem

$$\min_x f(x),$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable.

Gradient descent is a first-order optimization algorithm. To find a local minimum of a function using gradient descent, one takes steps proportional to the negative of the gradient of the function at the current point.

We can see that, if

$$x_1 = x_0 - \gamma \left((\nabla f)(x_0) \right)^T$$

for a small step size (learning rate) $\gamma > 0$ then $f(x_1) \leq f(x_0)$.

Iterate

$$x_{i+1} = x_i - \gamma_i \left((\nabla f)(x_i) \right)^T .$$

For suitable learning rate γ_i, the sequence $f(x_0) \geq f(x_1) \geq \cdots$ converges to a local minimum.

Example

Consider the problem of minimizing $f(x, y) = 2x^2 - 2xy + 1.5y^2$ using the gradient descent method.

We have

$$\nabla f(x, y) = \begin{bmatrix} 4x - 2y \\ -2x + 3y \end{bmatrix}$$

Example

Consider the problem of minimizing $f(x, y) = 2x^2 - 2xy + 1.5y^2$ using the gradient descent method.

We have

$$\nabla f(x, y) = \begin{bmatrix} 4x - 2y \\ -2x + 3y \end{bmatrix}$$

We will start from the point $(x_0, y_0) = (1, 2)$ and learning rate $\gamma = 0.05$.

Example

Consider the problem of minimizing $f(x,y) = 2x^2 - 2xy + 1.5y^2$ using the gradient descent method.

We have

$$\nabla f(x,y) = \begin{bmatrix} 4x - 2y \\ -2x + 3y \end{bmatrix}$$

We will start from the point $(x_0, y_0) = (1,2)$ and learning rate $\gamma = 0.05$. Iterative:

$$\begin{bmatrix} x_{i+1} \\ y_{i+1} \end{bmatrix} = \begin{bmatrix} x_i \\ y_i \end{bmatrix} - \gamma \begin{bmatrix} 4x_i - 2y_i \\ -2x_i + 3y_i \end{bmatrix} = \begin{bmatrix} 0.8x_i + 0.1y_i \\ 0.1x_i + 0.85y_i \end{bmatrix}$$

Example

Consider the problem of minimizing $f(x,y) = 2x^2 - 2xy + 1.5y^2$ using the gradient descent method.

We have

$$\nabla f(x,y) = \begin{bmatrix} 4x - 2y \\ -2x + 3y \end{bmatrix}$$

We will start from the point $(x_0, y_0) = (1, 2)$ and learning rate $\gamma = 0.05$.Iterative:

$$\begin{bmatrix} x_{i+1} \\ y_{i+1} \end{bmatrix} = \begin{bmatrix} x_i \\ y_i \end{bmatrix} - \gamma \begin{bmatrix} 4x_i - 2y_i \\ -2x_i + 3y_i \end{bmatrix} = \begin{bmatrix} 0.8x_i + 0.1y_i \\ 0.1x_i + 0.85y_i \end{bmatrix}$$

We obtain

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1.71 \end{bmatrix}, \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0.971 \\ 1.5535 \end{bmatrix}, \ldots$$

Example

Consider the problem of minimizing $f(x,y) = 2x^2 - 2xy + 1.5y^2$ using the gradient descent method.

We have

$$\nabla f(x,y) = \begin{bmatrix} 4x - 2y \\ -2x + 3y \end{bmatrix}$$

We will start from the point $(x_0, y_0) = (1,2)$ and learning rate $\gamma = 0.05$.Iterative:

$$\begin{bmatrix} x_{i+1} \\ y_{i+1} \end{bmatrix} = \begin{bmatrix} x_i \\ y_i \end{bmatrix} - \gamma \begin{bmatrix} 4x_i - 2y_i \\ -2x_i + 3y_i \end{bmatrix} = \begin{bmatrix} 0.8x_i + 0.1y_i \\ 0.1x_i + 0.85y_i \end{bmatrix}$$

We obtain

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1.71 \end{bmatrix}, \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0.971 \\ 1.5535 \end{bmatrix}, \ldots$$

Notice that the optimal solution is $(x,y) = (0,0)$.

Example

Consider the problem of minimizing $f(x, y) = 2x^2 - 2xy + 1.5y^2$ using the gradient descent method.

We have

$$\nabla f(x, y) = \begin{bmatrix} 4x - 2y \\ -2x + 3y \end{bmatrix}$$

We will start from the point $(x_0, y_0) = (1, 2)$ and learning rate $\gamma = 0.05$.Iterative:

$$\begin{bmatrix} x_{i+1} \\ y_{i+1} \end{bmatrix} = \begin{bmatrix} x_i \\ y_i \end{bmatrix} - \gamma \begin{bmatrix} 4x_i - 2y_i \\ -2x_i + 3y_i \end{bmatrix} = \begin{bmatrix} 0.8x_i + 0.1y_i \\ 0.1x_i + 0.85y_i \end{bmatrix}$$

We obtain

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1.71 \end{bmatrix}, \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0.971 \\ 1.5535 \end{bmatrix}, \dots$$

Notice that the optimal solution is $(x, y) = (0, 0)$.

Exercise. Write a python programe to find the local minima of the function

$$f(x) = x^2 - 2x + 8$$

starting from point $x_0 = 2$ by the gradient descent algorithm. Let us assume the learning rate 0.01 and the difference between x values from two consecutive iterations is less than 0.001.

Choosing a good step-size is important in gradient descent:

- If the step-size is too small, gradient descent can be slow.

Choosing a good step-size is important in gradient descent:

- If the step-size is too small, gradient descent can be slow.
- If the step-size is chosen too large, gradient descent can overshoot, fail to converge, or even diverge.

Choosing a good step-size is important in gradient descent:

- If the step-size is too small, gradient descent can be slow.
- If the step-size is chosen too large, gradient descent can overshoot, fail to converge, or even diverge.

Choosing a step size:

$$\gamma_i^\star = \operatorname*{argmin}_{\gamma} f\left(x_i - \gamma \nabla f(x_i)^T\right)$$

.

Choosing a good step-size is important in gradient descent:

- If the step-size is too small, gradient descent can be slow.
- If the step-size is chosen too large, gradient descent can overshoot, fail to converge, or even diverge.

Choosing a step size:

$$\gamma_i^\star = \underset{\gamma}{\operatorname{argmin}} f\left(x_i - \gamma \nabla f(x_i)^T\right)$$

.

Gradient descent with momentum

Remembers the update $\Delta x_i = x_i - x_{i-1}$ at each iteration i and determines the next update as a linear combination of the current and previous gradients:

$$x_{i+1} = x_i - \gamma_i \left((\nabla f)(x_i) \right)^T + \alpha \Delta x_i$$
$$\Delta x_i = x_i - x_{i-1} = \alpha \Delta x_{i-1} - \gamma_{i-1} \left((\nabla f)(x_{i-1}) \right)^T .$$

Consider the constrained optimization problem

$$\min f(x)$$
$$\text{subject to } g_i(x) \leq 0, i = 1, \ldots, m$$

We can convert the constrained problem into an unconstrained one:

$$J(x) = f(x) + \sum_{i=1}^{m} \mathbb{1}(g_i(x)),$$

Consider the constrained optimization problem

$$\min f(x)$$
$$\text{subject to } g_i(x) \leq 0, i = 1, \ldots, m$$

We can convert the constrained problem into an unconstrained one:

$$J(x) = f(x) + \sum_{i=1}^{m} \mathbb{1}(g_i(x)),$$

where $\mathbb{1}(z)$ is an infinite step function

$$\mathbb{1}(z) = \begin{cases} 0 & \text{if } z \leq 0 \\ \infty & \text{otherwise} \end{cases}.$$

FPT Education
FPT UNIVERSITY

Consider the constrained optimization problem

$$\min f(x)$$
$$\text{subject to } g_i(x) \leq 0, i = 1, \ldots, m$$

We can convert the constrained problem into an unconstrained one:

$$J(x) = f(x) + \sum_{i=1}^{m} \mathbb{1}(g_i(x)),$$

where $\mathbb{1}(z)$ is an infinite step function

$$\mathbb{1}(z) = \begin{cases} 0 & \text{if } z \leq 0 \\ \infty & \text{otherwise} \end{cases}.$$

The associated the Lagrangian

$$\mathcal{L}(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x)$$
$$= f(x) + \lambda^T g(x),$$

FPT Education FPT UNIVERSITY

Consider the constrained optimization problem

$$\min f(x)$$
$$\text{subject to } g_i(x) \leq 0, i = 1, \ldots, m$$

We can convert the constrained problem into an unconstrained one:

$$J(x) = f(x) + \sum_{i=1}^{m} \mathbb{1}(g_i(x)),$$

where $\mathbb{1}(z)$ is an infinite step function

$$\mathbb{1}(z) = \begin{cases} 0 & \text{if } z \leq 0 \\ \infty & \text{otherwise} \end{cases}.$$

The associated the Lagrangian

$$\begin{aligned} \mathcal{L}(x, \lambda) &= f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) \\ &= f(x) + \lambda^T g(x), \end{aligned}$$

Definition

The problem

$$\min f(x)$$
$$\text{subject to } g_i(x) \le 0, i = 1, \ldots, m$$

is known as the primal problem, corresponding to the primal variable $x \in \mathbb{R}^D$.

Definition

The problem

$$\min f(x)$$
$$\text{subject to } g_i(x) \leq 0, i = 1, \ldots, m$$

is known as the primal problem, corresponding to the primal variable $x \in \mathbb{R}^D$.

The associated Lagrangian dual problem is given by

$$\max_{\lambda \in (\mathbb{R}^+)^m} \mathcal{D}(\lambda) = \max_{\lambda \in (\mathbb{R}^+)^m} \min_{x \in \mathbb{R}^D} \mathcal{L}(x, \lambda),$$

Definition

The problem

$$\min f(x)$$
$$\text{subject to } g_i(x) \leq 0, i = 1, \ldots, m$$

is known as the primal problem, corresponding to the primal variable $x \in \mathbb{R}^D$.

The associated Lagrangian dual problem is given by

$$\max_{\lambda \in (\mathbb{R}^+)^m} \mathcal{D}(\lambda) = \max_{\lambda \in (\mathbb{R}^+)^m} \min_{x \in \mathbb{R}^D} \mathcal{L}(x, \lambda),$$

where $\lambda = [\lambda_1, \ldots, \lambda_m]$ are dual variables and $\mathcal{D}(\lambda) = \min_{x \in \mathbb{R}^D} \mathcal{L}(x, \lambda)$.

Definition

The problem

$$\min f(x)$$
$$\text{subject to } g_i(x) \leq 0, i = 1, \ldots, m$$

is known as the primal problem, corresponding to the primal variable $x \in \mathbb{R}^D$.

The associated Lagrangian dual problem is given by

$$\max_{\lambda \in (\mathbb{R}^+)^m} \mathcal{D}(\lambda) = \max_{\lambda \in (\mathbb{R}^+)^m} \min_{x \in \mathbb{R}^D} \mathcal{L}(x, \lambda),$$

where $\lambda = [\lambda_1, \ldots, \lambda_m]$ are dual variables and $\mathcal{D}(\lambda) = \min_{x \in \mathbb{R}^D} \mathcal{L}(x, \lambda)$.

Theorem (Minimax inequality)

For any function with two arguments $\varphi(x, y)$*:*

$$\max_y \min_x \varphi(x, y) \leq \min_x \max_y \varphi(x, y)$$

Theorem (Minimax inequality)

For any function with two arguments $\varphi(x, y)$:

$$\max_y \min_x \varphi(x, y) \le \min_x \max_y \varphi(x, y)$$

We can see that

$$J(x) = \max_{\lambda \in (\mathbb{R}^+)^m} \mathcal{L}(x, \lambda).$$

Theorem (Minimax inequality)

For any function with two arguments $\varphi(x, y)$:

$$\max_y \min_x \varphi(x, y) \leq \min_x \max_y \varphi(x, y)$$

We can see that

$$J(x) = \max_{\lambda \in (\mathbb{R}^+)^m} \mathcal{L}(x, \lambda).$$

Minimax inequality implies that

Theorem (Minimax inequality)

For any function with two arguments $\varphi(x,y)$:

$$\max_y \min_x \varphi(x,y) \le \min_x \max_y \varphi(x,y)$$

We can see that

$$J(x) = \max_{\lambda \in (\mathbb{R}^+)^m} \mathcal{L}(x,\lambda).$$

Minimax inequality implies that

Corollary (weak duality)

$$\min_{x \in \mathbb{R}^D} \max_{\lambda \in (\mathbb{R}^+)^m} \mathcal{L}(x,\lambda) \ge \max_{\lambda \in (\mathbb{R}^+)^m} \min_{x \in \mathbb{R}^D} \mathcal{L}(x,\lambda).$$

Theorem (Minimax inequality)

For any function with two arguments $\varphi(x, y)$:

$$\max_y \min_x \varphi(x, y) \le \min_x \max_y \varphi(x, y)$$

We can see that

$$J(x) = \max_{\lambda \in (\mathbb{R}^+)^m} \mathcal{L}(x, \lambda).$$

Minimax inequality implies that

Corollary (weak duality)

$$\min_{x \in \mathbb{R}^D} \max_{\lambda \in (\mathbb{R}^+)^m} \mathcal{L}(x, \lambda) \ge \max_{\lambda \in (\mathbb{R}^+)^m} \min_{x \in \mathbb{R}^D} \mathcal{L}(x, \lambda).$$

FPT Education
FPT UNIVERSITY

Definition

A set $\mathcal{C}$ is convex if for any $x, y \in C$ then the line segment joining x, y lie inside $\mathcal{C}$, i.e

$$\theta x + (1 - \theta) y \in \mathcal{C}.$$

Example

- The closed interval $[0, 8]$ is convex.

FPT Education
FPT UNIVERSITY

Definition

A set $\mathcal{C}$ is convex if for any $x, y \in C$ then the line segment joining x, y lie inside $\mathcal{C}$, i.e

$$\theta x + (1 - \theta) y \in \mathcal{C}.$$

Example

- The closed interval $[0, 8]$ is convex.
- The opened ball $B(a, r) := \{x \in \mathbb{R}^D : \|x - a\|_2 < r\}$ is convex.

FPT Education
FPT UNIVERSITY

Definition

A set $\mathcal{C}$ is convex if for any $x, y \in C$ then the line segment joining x, y lie inside $\mathcal{C}$, i.e

$$\theta x + (1 - \theta) y \in \mathcal{C}.$$

Example

- The closed interval $[0, 8]$ is convex.
- The opened ball $B(a, r) := \{x \in \mathbb{R}^D : \|x - a\|_2 < r\}$ is convex.
- The annulus $A(r, R) := \{x \in \mathbb{R}^2 : 0 < r_1 < \|x\|_2 < r_2\}$ is not convex.

FPT Education
FPT UNIVERSITY

Definition

A set $\mathcal{C}$ is convex if for any $x, y \in C$ then the line segment joining x, y lie inside $\mathcal{C}$, i.e

$$\theta x + (1 - \theta) y \in \mathcal{C}.$$

Example

- The closed interval $[0, 8]$ is convex.
- The opened ball $B(a, r) := \{x \in \mathbb{R}^D : \|x - a\|_2 < r\}$ is convex.
- The annulus $A(r, R) := \{x \in \mathbb{R}^2 : 0 < r_1 < \|x\|_2 < r_2\}$ is not convex.

FPT Education FPT UNIVERSITY

Definition

A set $\mathcal{C}$ is convex if for any $x, y \in C$ then the line segment joining x, y lie inside $\mathcal{C}$, i.e

$$\theta x + (1-\theta)y \in \mathcal{C}.$$

Example

- The closed interval $[0, 8]$ is convex.
- The opened ball $B(a, r) := \{x \in \mathbb{R}^D : \|x - a\|_2 < r\}$ is convex.
- The annulus $A(r, R) := \{x \in \mathbb{R}^2 : 0 < r_1 < \|x\|_2 < r_2\}$ is not convex.

Definition

Let $f:\mathbb{R}^D \to \mathbb{R}$ be a function whose domain is a convex set. The function f is a convex function if for all x, y in the domain of f, we have

$$f[\theta x + (1-\theta)y] \leq \theta f(x) + (1-\theta)f(y), \ \forall \ 0 \leq \theta \leq 1.$$

Definition

Let $f: \mathbb{R}^D \to \mathbb{R}$ be a function whose domain is a convex set. The function f is a convex function if for all x, y in the domain of f, we have

$$f[\theta x + (1-\theta)y] \leq \theta f(x) + (1-\theta) f(y), \ \forall\ 0 \leq \theta \leq 1.$$

Theorem

Suppose that $f: \mathbb{R}^D \to \mathbb{R}$ is differentiable. Then f is convex if and only if for any two points x, y it holds that

Definition

Let $f : \mathbb{R}^D \to \mathbb{R}$ be a function whose domain is a convex set. The function f is a convex function if for all x, y in the domain of f, we have

$$f[\theta x + (1-\theta)y] \leq \theta f(x) + (1-\theta) f(y), \ \forall \ 0 \leq \theta \leq 1.$$

Theorem

Suppose that $f : \mathbb{R}^D \to \mathbb{R}$ is differentiable. Then f is convex if and only if for any two points x, y it holds that

$$f(y) \geq f(x) + \nabla_x f(x)^T (y - x).$$

Definition

Let $f : \mathbb{R}^D \to \mathbb{R}$ be a function whose domain is a convex set. The function f is a convex function if for all x, y in the domain of f, we have

$$f[\theta x + (1-\theta)y] \leq \theta f(x) + (1-\theta) f(y), \;\; \forall\; 0 \leq \theta \leq 1.$$

Theorem

Suppose that $f : \mathbb{R}^D \to \mathbb{R}$ is differentiable. Then f is convex if and only if for any two points x, y it holds that

$$f(y) \geq f(x) + \nabla_x f(x)^T (y - x).$$

Furthermore, if $f(x)$ is twice differentiable then f is convex if and only if its Hessian $\nabla_x^2 f(x)$ is positive semi-definite.

Definition

Let $f : \mathbb{R}^D \to \mathbb{R}$ be a function whose domain is a convex set. The function f is a convex function if for all x, y in the domain of f, we have

$$f[\theta x + (1-\theta)y] \leq \theta f(x) + (1-\theta) f(y), \ \forall \ 0 \leq \theta \leq 1.$$

Theorem

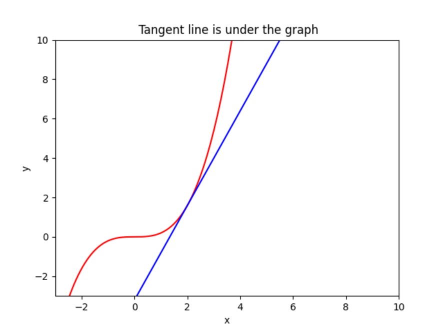
Suppose that $f : \mathbb{R}^D \to \mathbb{R}$ is differentiable. Then f is convex if and only if for any two points x, y it holds that

$$f(y) \geq f(x) + \nabla_x f(x)^T (y - x).$$

Furthermore, if $f(x)$ is twice differentiable then f is convex if and only if its Hessian $\nabla_x^2 f(x)$ is positive semi-definite.

Example

The function $f(x) = x \ln x$ is convex on $x > 0$.

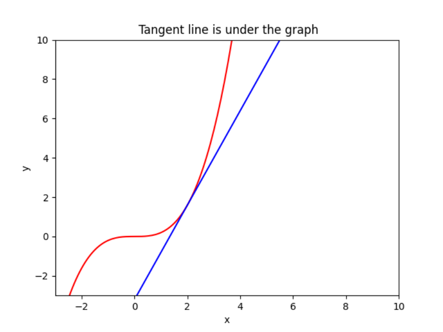


Example

The function $f(x, y) = x^2 + y^2$ is convex on $\mathbb{R}^2$

Example

The function $f(x) = x \ln x$ is convex on $x > 0$.



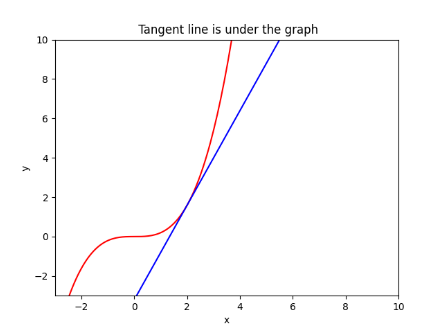
Example

The function $f(x, y) = x^2 + y^2$ is convex on $\mathbb{R}^2$since its Hessian

$$\nabla^2_{x,y} f(x, y) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \quad \text{is positive definite.}$$

Example

The function $f(x) = x \ln x$ is convex on $x > 0$.



Example

The function $f(x, y) = x^2 + y^2$ is convex on $\mathbb{R}^2$since its Hessian

$$\nabla^2_{x,y} f(x, y) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \quad \text{is positive definite.}$$

Definition

A constrained optimization problem is called a convex optimization problem if

$$\begin{aligned} &\min_x f(x) \\ \text{subject to } &g_i(x) \le 0, \text{ for all } i = 1, \dots, m \\ &h_j(x) = 0, \text{ for all } j = 1, \dots, n \end{aligned}$$

where functions $f(x), g_i(x)$ are convex functions and $\{x : h_j(x) = 0\}$ are convex sets.

Definition

A constrained optimization problem is called a convex optimization problem if

$$\begin{aligned} &\min_x f(x) \\ \text{subject to } & g_i(x) \leq 0, \ \text{for all } i = 1, \ldots, m \\ & h_j(x) = 0, \ \text{for all } j = 1, \ldots, n \end{aligned}$$

where functions $f(x), g_i(x)$ are convex functions and $\{x : h_j(x) = 0\}$ are convex sets.

Definition

A constrained optimization problem is called a convex optimization problem if

$$\begin{aligned} &\min_x f(x) \\ \text{subject to } &g_i(x) \leq 0, \text{ for all } i = 1, \dots, m \\ &h_j(x) = 0, \text{ for all } j = 1, \dots, n \end{aligned}$$

where functions $f(x), g_i(x)$ are convex functions and $\{x : h_j(x) = 0\}$ are convex sets.

7.3.1 Linear programming

We consider a special cases of convex optimization problem called linear program

$$\min_{x \in \mathbb{R}^n} c^T x \tag{1}$$

$$\text{subject to } Ax \le b$$

where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$. The Lagrangian is given by

$$\begin{aligned}\mathcal{L}(x, \lambda) &= c^T x + \lambda^T (Ax - b) \\ &= \left(c^T + A^T \lambda\right)^T x - \lambda^T b.\end{aligned}$$

We consider a special cases of convex optimization problem called linear program

$$\min_{x \in \mathbb{R}^n} c^T x \tag{1}$$
$$\text{subject to } Ax \leq b$$

where $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m, c \in \mathbb{R}^n$.The Lagrangian is given by

$$\begin{aligned}\mathcal{L}(x, \lambda) &= c^T x + \lambda^T (Ax - b) \\ &= \left(c^T + A^T \lambda\right)^T x - \lambda^T b.\end{aligned}$$

We have

$$\frac{\partial \mathcal{L}(x, \lambda)}{\partial x} = c^T + A^T \lambda = 0.$$

FPT Education FPT UNIVERSITY

We consider a special cases of convex optimization problem called linear program

$$\min_{x \in \mathbb{R}^n} c^T x \tag{1}$$

$$\text{subject to } Ax \leq b$$

where $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m, c \in \mathbb{R}^n$.The Lagrangian is given by

$$\begin{aligned}\mathcal{L}(x, \lambda) &= c^T x + \lambda^T (Ax - b) \\ &= \left(c^T + A^T \lambda\right)^T x - \lambda^T b.\end{aligned}$$

We have

$$\frac{\partial \mathcal{L}(x, \lambda)}{\partial x} = c^T + A^T \lambda = 0.$$

Hence the dual Lagrangian

$$\mathcal{D}(\lambda) = -\lambda^T b.$$

7.3.1 Linear programming

We consider a special cases of convex optimization problem called linear program

$$\min_{x \in \mathbb{R}^n} c^T x \tag{1}$$
$$\text{subject to } Ax \leq b$$

where $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m, c \in \mathbb{R}^n$.The Lagrangian is given by

$$\begin{aligned}\mathcal{L}(x, \lambda) &= c^T x + \lambda^T (Ax - b) \\ &= \left(c^T + A^T \lambda\right)^T x - \lambda^T b.\end{aligned}$$

We have

$$\frac{\partial \mathcal{L}(x, \lambda)}{\partial x} = c^T + A^T \lambda = 0.$$

Hence the dual Lagrangian

$$\mathcal{D}(\lambda) = -\lambda^T b.$$

The dual optimization

$$\max_{\lambda \in (\mathbb{R}^+)^m} -b^T \lambda \quad (2)$$

$$\text{subject to } c + A^T \lambda = 0.$$

This is also a linear program, but with m variables. We can solve the primal (1) or the dual (2) program depending on whether m or n is larger.

The dual optimization

$$\max_{\lambda \in (\mathbb{R}^+)^m} -b^T \lambda \tag{2}$$
$$\text{subject to } c + A^T \lambda = 0.$$

This is also a linear program, but with m variables. We can solve the primal (1) or the dual (2) program depending on whether m or n is larger.

7.3.2 Quadratic Programming

Consider a quadratic program:

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} x^T Q x + c^T x \qquad (3)$$
$$\text{subject to } Ax \leq b$$

where $Q \in \mathbb{R}^{n \times n}$ is positive definite, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^n$ and $c \in \mathbb{R}^n$.

7.3.2 Quadratic Programming

Consider a quadratic program:

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} x^T Q x + c^T x \tag{3}$$
$$\text{subject to } Ax \le b$$

where $Q \in \mathbb{R}^{n \times n}$ is positive definite, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^n$ and $c \in \mathbb{R}^n$.

The Lagrange is given by

$$\begin{aligned} \mathcal{L}(x, \lambda) &= \frac{1}{2} x^T Q x + c^T x + \lambda^T (Ax - b) \\ &= \frac{1}{2} x^T Q x + (c + A^T \lambda)^T x - \lambda^T b. \end{aligned}$$

7.3.2 Quadratic Programming

Consider a quadratic program:

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} x^T Q x + c^T x \quad (3)$$
$$\text{subject to } Ax \leq b$$

where $Q \in \mathbb{R}^{n \times n}$ is positive definite, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^n$ and $c \in \mathbb{R}^n$.

The Lagrange is given by

$$\begin{aligned} \mathcal{L}(x, \lambda) &= \frac{1}{2} x^T Q x + c^T x + \lambda^T (Ax - b) \\ &= \frac{1}{2} x^T Q x + (c + A^T \lambda)^T x - \lambda^T b. \end{aligned}$$

The partial derivative

$$\frac{\mathcal{L}(x,\lambda)}{\partial x} = Qx + (c + A^T\lambda) = 0 \Rightarrow x = -Q^{-1}(c + A^T\lambda)$$

We get the dual Lagrangian

$$\mathcal{D}(\lambda) = -\frac{1}{2}(c + A^T\lambda)^T Q^{-1}(c + A^T\lambda) - \lambda^T b.$$

The partial derivative

$$\frac{\mathcal{L}(x,\lambda)}{\partial x} = Qx + (c + A^T\lambda) = 0 \Rightarrow x = -Q^{-1}(c + A^T\lambda)$$

We get the dual Lagrangian

$$\mathcal{D}(\lambda) = -\frac{1}{2}(c + A^T\lambda)^T Q^{-1}(c + A^T\lambda) - \lambda^T b.$$

Therefore, the dual optimization problem is given by

$$\max_{\lambda\in(\mathbb{R}^+)^m} -\frac{1}{2}(c + A^T\lambda)^T Q^{-1}(c + A^T\lambda) - \lambda^T b$$

subject to $\lambda \geq 0$.

The partial derivative

$$\frac{\mathcal{L}(x,\lambda)}{\partial x} = Qx + (c + A^T\lambda) = 0 \Rightarrow x = -Q^{-1}(c + A^T\lambda)$$

We get the dual Lagrangian

$$\mathcal{D}(\lambda) = -\frac{1}{2}(c + A^T\lambda)^T Q^{-1}(c + A^T\lambda) - \lambda^T b.$$

Therefore, the dual optimization problem is given by

$$\max_{\lambda \in (\mathbb{R}^+)^m} -\frac{1}{2}(c + A^T\lambda)^T Q^{-1}(c + A^T\lambda) - \lambda^T b$$

subject to $\lambda \geq 0$.

Summary

We have studied:

- optimization using gradient method;
- constrained optimization;
- convex optimization.

Exercises for practice: 7.1-7.8 (page 247, 248).