Optimality conditions. Optimization with equality / inequality conditions. KKT.

Seminar

Optimization for ML. Faculty of Computer Science. HSE University

Optimality Conditions. Important notions recap

$$f(x) \to \min_{x \in S}$$

A set S is usually called a budget set.

- A point x^* is a global minimizer if $f(x^*) \leq f(x)$ for all x.
- A point x^* is a local minimizer if there exists a neighborhood N of x^* such that $f(x^*) \leq f(x)$ for all $x \in N$.
- A point x^* is a strict local minimizer (also called a strong local minimizer) if there exists a neighborhood N of x^* such that $f(x^*) < f(x)$ for all $x \in N$ with $x \neq x^*$.
- We call x^* a stationary point (or critical) if $\nabla f(x^*) = 0$. Any local minimizer must be a stationary point.

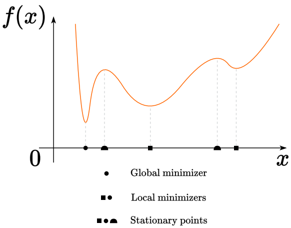

Figure 1: Illustration of different stationary (critical) points

Unconstrained optimization recap

First-Order Necessary Conditions

If x^* is a local minimizer and f is continuously differentiable in an open neighborhood, then

$$\nabla f(x^*) = 0 \tag{1}$$

Second-Order Sufficient Conditions

Suppose that $\nabla^2 f$ is continuous in an open neighborhood of x^* and that

$$\nabla f(x^*) = 0 \quad \nabla^2 f(x^*) \succ 0. \tag{2}$$

Then x^* is a strict local minimizer of f.

Optimization with equality conditions

Consider simple yet practical case of equality constraints:

$$
\begin{split}
& f(x) \to \min\limits_{x \in \mathbb{R}^n} \\
\text{s.t. } & h_i(x) = 0, i = 1, \ldots, p
\end{split}
$$

Lagrange multipliers recap

The basic idea of Lagrange method implies the switch from conditional to unconditional optimization through increasing the dimensionality of the problem:

$$L(x, \nu) = f(x) + \sum_{i=1}^{p} \nu_i h_i(x) = f(x) + \nu^T h(x) \to \min_{x \in \mathbb{R}^n, \nu \in \mathbb{R}^p}$$

Lagrange multipliers recap

The basic idea of Lagrange method implies the switch from conditional to unconditional optimization through increasing the dimensionality of the problem:

$$L(x, \nu) = f(x) + \sum_{i=1}^{p} \nu_i h_i(x) = f(x) + \nu^T h(x) \to \min_{x \in \mathbb{R}^n, \nu \in \mathbb{R}^p}$$

Necessery conditions:

$$\nabla_x L(x^*, \nu^*) = 0$$

$$\nabla_\nu L(x^*, \nu^*) = 0$$

Sufficient conditions:

$$\langle y, \nabla^2_{xx} L(x^*, \nu^*) y \rangle > 0,$$

$$\forall y \neq 0 \in \mathbb{R}^n : \nabla h_i(x^*)^T y = 0$$

Optimization with inequality conditions

Consider simple yet practical case of inequality constraints:

$$
\begin{split}
& f(x) \to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } & g(x) \leq 0
\end{split}
$$

Optimization with inequality conditions

Consider simple yet practical case of inequality constraints:

$$
f(x) \to \min_{x \in \mathbb{R}^n}
$$
$$
\text{s.t. } g(x) \leq 0
$$

$g(x) \leq 0$ **is inactive**. $g(x^*) < 0$:

$$
g(x^*) < 0
$$
$$
\nabla f(x^*) = 0
$$
$$
\nabla^2 f(x^*) > 0
$$

$g(x) \leq 0$ **is active**. $g(x^*) = 0$:

$$
g(x^*) = 0
$$
$$
-\nabla f(x^*) = \lambda \nabla g(x^*), \lambda > 0
$$
$$
\langle y, \nabla^2_{xx} L(x^*, \lambda^*) y \rangle > 0,
$$
$$
\forall y \neq 0 \in \mathbb{R}^n : \nabla g(x^*)^\top y = 0
$$

General formulation

General problem of mathematical programming:

$$
\begin{split}
& f_0(x) \to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } & f_i(x) \leq 0, \; i = 1, \ldots, m \\
& h_i(x) = 0, \; i = 1, \ldots, p
\end{split}
$$

General formulation

General problem of mathematical programming:

$$
\begin{split}
& f_0(x) \to \min_{x \in \mathbb{R}^n}\\
\text{s.t. } & f_i(x) \leq 0, \; i = 1, \ldots, m\\
& h_i(x) = 0, \; i = 1, \ldots, p
\end{split}
$$

The solution involves constructing a Lagrange function:

$$
L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x)
$$

KKT Necessary conditions

Let x^*, (λ^*, ν^*) be a solution to a mathematical programming problem with zero duality gap (the optimal value for the primal problem p^* is equal to the optimal value for the dual problem d^*). Let also the functions f_0, f_i, h_i be differentiable.

KKT Necessary conditions

Let x^*, (λ^*, ν^*) be a solution to a mathematical programming problem with zero duality gap (the optimal value for the primal problem p^* is equal to the optimal value for the dual problem d^*). Let also the functions f_0, f_i, h_i be differentiable.

$$
\begin{aligned}
&(1)\nabla_x L(x^*, \lambda^*, \nu^*) = 0 \\
&(2)\nabla_\nu L(x^*, \lambda^*, \nu^*) = 0 \\
&(3)\lambda_i^* \geq 0, i = 1, \ldots, m \\
&(4)\lambda_i^* f_i(x^*) = 0, i = 1, \ldots, m \\
&(5)f_i(x^*) \leq 0, i = 1, \ldots, m
\end{aligned}
$$

KKT Some regularity conditions

These conditions are needed in order to make KKT solutions the necessary conditions. Some of them even turn necessary conditions into sufficient. For example, Slater's condition:

KKT Some regularity conditions

These conditions are needed in order to make KKT solutions the necessary conditions. Some of them even turn necessary conditions into sufficient. For example, Slater's condition:

> If for a convex problem (i.e., assuming minimization, f_0, f_i are convex and h_i are affine), there exists a point x such that $h(x) = 0$ and $f_i(x) < 0$ (existance of a strictly feasible point), then we have a zero duality gap and KKT conditions become necessary and sufficient.

KKT Sufficient conditions

For smooth, non-linear optimization problems, a second order sufficient condition is given as follows. The solution x^*, λ^*, ν^*, which satisfies the KKT conditions (above) is a constrained local minimum if for the Lagrangian,

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x)$$

the following conditions hold:

$$\begin{aligned}
&\langle y, \nabla^2_{xx} L(x^*, \lambda^*, \nu^*) y \rangle > 0 \\
&\forall y \neq 0 \in \mathbb{R}^n : \nabla h_i(x^*)^\top y = 0, \nabla f_0(x^*)^\top y \leq 0, \nabla f_j(x^*)^\top y = 0 \\
&i = 1, \ldots, p \quad \forall j : f_j(x^*) = 0
\end{aligned}$$

Problem 1

Question

Function $f : E \to \mathbb{R}$ is defined as

$$f(x) = \ln\left(-Q(x)\right)$$

where $E = \{x \in \mathbb{R}^n : Q(x) < 0\}$ and

$$Q(x) = \frac{1}{2} x^\top A x + b^\top x + c$$

with $A \in \mathbb{S}^n_{++}$, $b \in \mathbb{R}^n$, $c \in \mathbb{R}$.
Find the maximizer x^* of the function f.

Problem 2

Question

Give an explicit solution of the following task.

$$
\begin{aligned}
& f(x, y) = x + y \to \min \\
\text{s.t. } & x^2 + y^2 = 1
\end{aligned}
$$

where $x, y \in \mathbb{R}$.

Problem 3

Question

Give an explicit solution of the following task.

$$\langle c, x \rangle + \sum_{i=1}^n x_i \log x_i \to \min_{x \in \mathbb{R}^n}$$

$$\text{s.t. } \sum_{i=1}^n x_i = 1,$$

where $x \in \mathbb{R}^n_{++}, c \neq 0$.

Problem 4

Question

Let $e_1, \ldots, e_n$ be a standart basis in $\mathbb{R}^n$. Show that:

$$\max_{X \in \mathbb{S}^n_{++}} \det(X) : ||Xe_i|| \leq 1 \forall i \in 1, \ldots, n$$

Has a unique solution I_n, and derive the Hadamard inequality:

$$\det(X) \leq \prod_{i=1}^{n} ||Xe_i|| \forall X \in \mathbb{S}^n_{++}$$

Problem 5

Question

Given $y \in \{-1, 1\}$, and $X \in \mathbb{R}^{n \times p}$, the Support Vector Machine problem is:

$$
\begin{split}
&\frac{1}{2}||w||_2^2 + C\sum_{i=1}^n \xi_i \to \min_{w,w_0,\xi_i} \\
\text{s.t. } & \xi_i \geq 0, i = 1, \dots, n \\
& y_i(x_i^T w + w_0) \geq 1 - \xi_i, i = 1, \dots, n
\end{split}
$$

find the KKT stationarity condition.

Problem 6 (BONUS)

For some $\Sigma, \Sigma_0 \in \mathbb{S}^n_{++}$ define a KL Divergence between two Gaussian distributions as:

$$D(\Sigma, \Sigma_0) = \frac{1}{2}(\langle \Sigma_0^{-1}, \Sigma \rangle - \log\det(\Sigma_0^{-1}\Sigma) - n)$$

Now let $H \in \mathbb{S}^n_{++}$ and $y, x \in \mathbb{R}^n$: $\langle y, s \rangle > 0$

We would like to solve the following constrained minimization task.

$$\min_{X \in \mathbb{S}^n_{++}} \{D(X^{-1}, H^{-1}) | Xy = s\}$$

Prove that it hass a unique sollution and it is equal to:

$$(I_n - \frac{sy^T}{y^Ts})H(I_n - \frac{ys^T}{y^Ts}) + \frac{ss^T}{y^Ts}$$

Adversarial Attacks

Definition: Adversarial attacks are techniques used to fool DL models by adding small perturbations to the input data.

Context: These perturbations are often imperceptible to humans but can significantly affect model performance.

Optimization Perspective: We can frame adversarial attacks as a constrained optimization problem where the goal is to minimize/maximize the loss function while keeping the perturbation within a certain limit (norm constraint).

One of the most simple yet in some ways efficient methods is Fast Gradient Sign Method (FGSM).

The Fast Gradient Sign Method (FGSM) is an adversarial attack technique that generates adversarial examples by applying a small perturbation in the direction of the gradient of the loss function. Fromally:

Given x we would like to build and adversarial example such that it is not far from the initial point:

$$x' = x + \varepsilon \cdot \text{sgn}(\nabla_x L(x, y)), \text{s.t. } ||x - x'|| \leq \varepsilon$$

So in a nutshell we perfrom a gradient ascent on an image (== maximizing loss w.r.t to that image).

#TODO: Add picture here.