Optimization Methods (CS1.404), Spring 2024
Lecture 20

Naresh Manwani

Machine Learning Lab, IIIT-H

March 21st, 2024

KKT Optimality Conditions: First Order

KKT Optimality Conditions of First Order

Consider the problem $\min\ f(\mathbf{x})$ such that $h_j(\mathbf{x}) \leq 0,\ j = 1 \dots l,\ \mathbf{x} \in \mathbb{R}^n$. Assume that $\mathbf{x}^* \in \mathcal{X}$ to be a regular point and $\mathbf{x}^*$ is a local minima. Then there exist $\lambda_j\ j = 1 \dots l$, such that

$$\nabla f(\mathbf{x}^*) + \sum_{j=1}^{l} \lambda_j \nabla h_j(\mathbf{x}^*) = \mathbf{0}$$

$$\lambda_j h_j(\mathbf{x}^*) = 0; \quad j = 1, \dots, l$$

$$\lambda_j \geq 0; \quad j = 1, \dots, l$$

- These are first order KKT necessary conditions.
- KKT point: $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$, where $\boldsymbol{\lambda}^* = [\lambda_1^* \ \ \lambda_2^* \ \ \dots \ \ \lambda_l^*]^T$.

Lagrangian Function

- Lagrangian function is represented as:

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = \nabla f(\mathbf{x}) + \sum_{j=1}^{l} \lambda_j \nabla h_j(\mathbf{x})$$

- KKT Conditions imply

$$\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*) = 0$$
$$\lambda_j^* \geq 0; \;\; j = 1 \ldots l \quad (\lambda' s \text{ are called Lagrange multipliers.})$$
$$\lambda_j^* h_j(\mathbf{x}^*) = 0; \;\; j = 1 \ldots l \quad (\text{Complementary slackness conditions.})$$
$$\lambda_j^* = 0; \;\; \forall j \in \mathcal{A}(\mathbf{x}^*)$$

- Note that for active constraints, $\lambda_j^* h_j(\mathbf{x}^*) = 0$ because $h_j(\mathbf{x}^*) = 0$. Thus, λ_j^* can be zero or greater than zero.
- For non-active constraints, $h_j(\mathbf{x}^*) < 0$. Thus, $\lambda_j^* h_j(\mathbf{x}^*) = 0$ implies $\lambda_j^* = 0$.

Necessity of the KKT Conditions Under Regularity Condition for Convex Optimization Problem

Theorem

Let $\mathbf{x}^*$ be a regular point and is an optimal solution of the problem

$$
\begin{aligned}
CP: \quad & \min_{\mathbf{x} \in \mathbb{R}^n} \; f(\mathbf{x}) \\
& s.t. \;\; h_j(\mathbf{x}) \leq 0, \; j = 1 \ldots l
\end{aligned}
$$

where $f(\mathbf{x})$ and $h_1(\mathbf{x}), \ldots, h_l(\mathbf{x})$ are continuously differentiable convex functions over $\mathbb{R}^n$. Then, there exists multipliers $\lambda_1, \ldots, \lambda_l \geq 0$, such that

$$
\nabla f(\mathbf{x}^*) + \sum_{j=1}^{l} \lambda_j \nabla h_j(\mathbf{x}^*) = \mathbf{0}
$$

$$
\lambda_j h_j(\mathbf{x}^*) = 0; \;\; j = 1 \ldots l.
$$

Sufficiency of the KKT conditions Under Regularity Condition for Convex Optimization Problems

- KKT conditions are necessary optimality conditions under the regularity condition.
- When the problem is convex, the KKT conditions are always sufficient and no further condition is required.

Theorem

Consider the convex optimization problem:

$$CP: \quad \min_{\mathbf{x}\in\mathbb{R}^n} \; f(\mathbf{x})$$
$$s.t. \;\; h_j(\mathbf{x}) \leq 0, \; j = 1 \ldots l$$

where $f(\mathbf{x}), h_1(\mathbf{x}), \ldots, h_l(\mathbf{x})$ are continuously differentiable convex functions over $\mathbb{R}^n$. Let there exist multipliers $\lambda_1, \ldots, \lambda_l \geq 0$ such that

$$\nabla f(\mathbf{x}^*) + \sum_{j=1}^{l} \lambda_j \nabla h_j(\mathbf{x}^*) = \mathbf{0}$$
$$\lambda_j h_j(\mathbf{x}^*) = 0; \;\; j = 1 \ldots l.$$

Then, $\mathbf{x}^*$ is an optimal solution.

Slater's Condition

Slater's Condition

Let $h_j(\mathbf{x}) \leq 0;\ j = 1 \ldots l$ are convex inequalities such that $h_j(\mathbf{x}^*),\ j = 1 \ldots l$ are convex functions. Slater's condition is satisfied for these inequalities if there exists a point $\hat{\mathbf{x}}$ such that

$$h_j(\hat{\mathbf{x}}) < 0;\ j = 1 \ldots l.$$

Thus, Slater's condition requires that there exists a point that strictly satisfies the constraints. In other words, the interior of the feasible set is non-empty.

- Slater's condition does not require, like in the regularity condition, an apriori knowledge on the point that is a candidate to be an optimal solution.
- Checking the validity of Slater's condition is much easier task than checking regularity.

Necessity of the KKT Conditions Under Slater's Condition for Convex Optimization Problem

Theorem

Let $\mathbf{x}^*$ be an optimal solution of the problem

$$CP: \quad \min_{\mathbf{x}\in\mathbb{R}^n} \; f(\mathbf{x})$$
$$s.t. \; h_j(\mathbf{x}) \leq 0, \; j = 1 \dots l$$

where $f(\mathbf{x})$ and $h_1(\mathbf{x}), \dots, h_l(\mathbf{x})$ are continuously differentiable convex functions over $\mathbb{R}^n$. In addition, suppose there exists a point $\hat{\mathbf{x}}$ such that

$$h_j(\hat{\mathbf{x}}) < 0; \; j = 1 \dots l.$$

Then, there exists multipliers $\lambda_1, \dots, \lambda_l \geq 0$, such that

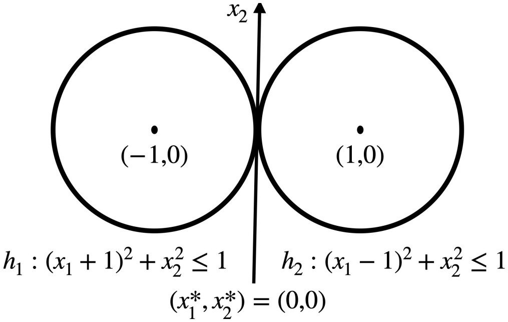
$$\nabla f(\mathbf{x}^*) + \sum_{j=1}^{l} \lambda_j \nabla h_j(\mathbf{x}^*) = \mathbf{0}$$
$$\lambda_j h_j(\mathbf{x}^*) = 0; \;\; j = 1 \dots l.$$

Not all Problems Satisfy Slater's Condition

Consider the optimization problem as follows.

$$\begin{aligned} \min \;\; & x_1 + x_2 \\ & (x_1+1)^2 + x_2^2 \le 1 \\ & (x_1-1)^2 + x_2^2 \le 1 \end{aligned}$$

Here, Feasible set $\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^2 \mid (x_1+1)^2 + x_2^2 \le 1,\ (x_1-1)^2 + x_2^2 \le 1\} = \{(0,0)\}$. At this point, both the constraints are satisfied with equality. Thus, it does not satisfy Slater's condition.



Equality Constraint Problems

Optimization problem with equality constraints is given as below.

$$\min_{\mathbf{x}\in\mathbb{R}^n} \; f(\mathbf{x})$$

$$s.t. \;\; e_i(\mathbf{x}) = 0; \; i = 1 \dots m$$

where $f(\mathbf{x})$, $e_1(\mathbf{x}), \dots, e_m(\mathbf{x})$ are smooth functions over $\mathbb{R}^n$.

Regular Point for Equality Constraint Problems

Definition

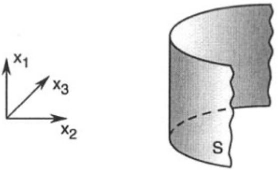
A point $\mathbf{x}^*$ satisfying the equality constraints $e_1(\mathbf{x}^*) = 0, \ldots, e_m(\mathbf{x}^*) = 0$ is said to be a regular point of the constraints if the gradient vectors $\nabla e_1(\mathbf{x}^*), \ldots, \nabla e_m(\mathbf{x}^*)$ are linearly independent. Let $D\mathbf{e}(\mathbf{x}^*)$ be the Jacobian matrix of $\mathbf{e} = [e_1, \ldots, e_m]^T$ at $\mathbf{x}^*$, given by

$$D\mathbf{e}(\mathbf{x}^*) = \begin{bmatrix} De_1(\mathbf{x}^*) \\ \vdots \\ e_m(\mathbf{x}^*) \end{bmatrix} = \begin{bmatrix} \nabla e_1(\mathbf{x}^*)^T \\ \vdots \\ \nabla e_m(\mathbf{x}^*)^T \end{bmatrix}$$

Then, $\mathbf{x}^*$ is regular if and only if rank $D\mathbf{e}(\mathbf{x}^*) = m$. That is, the Jacobian matrix is of full rank.

Example 1

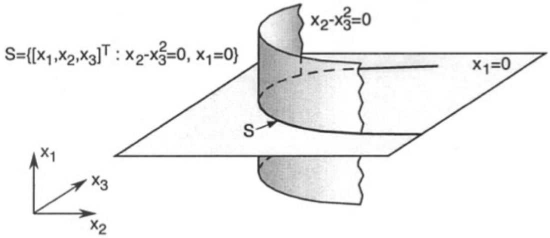
- Let there is a single equality constraint in $\mathbb{R}^3$. Thus, $n = 3, m = 1$.
- $e(x_1, x_2, x_3) = x_2 - x_3^2 = 0$
- $\nabla e(x_1, x_2, x_3) = [0, 1, -2x_3]^T$. Hence, for any (x_1, x_2, x_3), $\nabla e(x_1, x_2, x_3) \neq \mathbf{0}$.
- In this case, $Dim(S) = \dim\{(x_1, x_2, x_3) \mid \nabla e(x_1, x_2, x_3) = \mathbf{0}\} = n - m = 2$



$S=\{[x_1,x_2,x_3]^T : x_2-x_3^2=0\}$

Example 2

- Let there are two equalities constraint in $\mathbb{R}^3$. Thus, $n = 3, m = 2$.
- $e_1(x_1, x_2, x_3) = x_1 = 0$ and $e_2(x_1, x_2, x_3) = x_2 - x_3^2 = 0$
- $\nabla e_1(x_1, x_2, x_3) = [0, 0, 1]^T$ and $\nabla e_2(x_1, x_2, x_3) = [0, 1, -2x_3]^T$. Hence, the vectors $\nabla e_1(x_1, x_2, x_3)$ and $\nabla e_2(x_1, x_2, x_3)$ are linearly independent in $\mathbb{R}^3$.
- In this case, $Dim(S) = \dim\{(x_1, x_2, x_3) \mid \nabla e_1(x_1, x_2, x_3) = \mathbf{0}, \nabla e_2(x_1, x_2, x_3) = \mathbf{0}\} = n - m = 1$.



Dimension of Feasible Set of Set of Equality Constraints

The set of equality constraints $e_1(\mathbf{x}^*) = 0, \ldots, e_m(\mathbf{x}^*) = 0$, $e_i : \mathbb{R}^n \to \mathbb{R}$, describes a surface

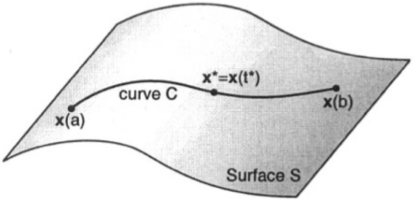
$$S = \{\mathbf{x} \in \mathbb{R}^n \mid e_1(\mathbf{x}^*) = 0, \ldots, e_m(\mathbf{x}^*) = 0\}.$$

Assuming the point in S are regular, the dimension of the surface S is $n - m$.

Curve on the Surface

Definition

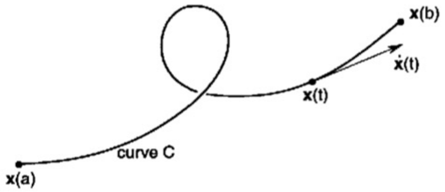
A curve C on a surface S is a set of points $\{\mathbf{x}(t) \in S \mid t \in (a, b)\}$, continuously parameterized by $t \in (a, b)$, that is, $\mathbf{x} : (a, b) \to S$ is a continuous function.



- All the points on the curve satisfy the equation describing the surface.
- The curve passes through the point $\mathbf{x}^*$ if there exist $t^* \in (a, b)$ such that $\mathbf{x}(t^*) = \mathbf{x}^*$.

Curve on the Surface

- The curve $C = \{\mathbf{x}(t) \in S \mid t \in (a, b)\}$ is differentiable if $\mathbf{x}'(t) = \frac{\partial \mathbf{x}(t)}{\partial t} = \begin{bmatrix} x_1'(t) \\ \vdots \\ x_n'(t) \end{bmatrix}$ exists for all $t \in (a, b)$.
- The curve $C = \{\mathbf{x}(t) \in S \mid t \in (a, b)\}$ is twice-differentiable if $\mathbf{x}''(t) = \frac{\partial^2 \mathbf{x}(t)}{\partial t^2} = \begin{bmatrix} x_1''(t) \\ \vdots \\ x_n''(t) \end{bmatrix}$ exists for all $t \in (a, b)$.
- The vector $\mathbf{x}'(t)$ is the direction of the tangent to the curve at $\mathbf{x}(t)$.



Gradient is perpendicular to the level curve

Theorem

Consider a function $\mathbf{e} : \mathbb{R}^n \to \mathbb{R}^m$ and $\mathbf{e} \in \mathbb{C}^1$. Consider the level set

$$S = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{e}(\mathbf{x}) = \mathbf{0}\}$$

Then, for any point $\mathbf{x}_0$ in S, the Jacobian $\nabla \mathbf{e}(\mathbf{x}_0)$ is perpendicular to S.

Proof:

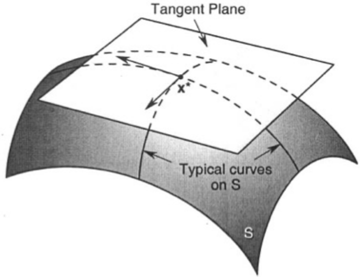
- We need to show that for any vector $\mathbf{a}$, which is tangent to S at $\mathbf{x}_0$, we have that $\mathbf{a}$ is perpendicular to $\nabla \mathbf{e}(\mathbf{x}_0)$.
- If $\mathbf{a}$ is tangent to S, we can find a parametrized curve $\mathbf{x}(t)$ lying in S such that $\mathbf{x}_0 = \mathbf{x}(t_0)$ and $\mathbf{x}'(t_0) = \mathbf{a}$.

Tangent Space

Definition

Tangent space at a point $\mathbf{x}^*$ on the surface $S = \{\mathbf{x} \in \mathbb{R}^n \mid e_1(\mathbf{x}^*) = 0, \ldots, e_m(\mathbf{x}^*) = 0\}$ is the set

$$\begin{aligned} T(\mathbf{x}^*) &= \{\mathbf{d} \mid D\mathbf{e}(\mathbf{x}^*)\mathbf{d} = \mathbf{0}\} \\ &= \{\mathbf{d} \mid \nabla e_1(\mathbf{x}^*)^T \mathbf{d} = 0, \ldots, \nabla e_m(\mathbf{x}^*)^T \mathbf{d} = 0\} \end{aligned}$$



- Tangent space at $\mathbf{x}^*$ is the null-space of $D\mathbf{e}(\mathbf{x}^*)$, which is a subspace of $\mathbb{R}^n$.
- Assuming $\mathbf{x}^*$ is a regular point, dimension of the tangent space $T(\mathbf{x}^*)$ is $n - m$.
- Tangent space passes through the origin.

Example of Tangent Space

- Let $S = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid e_1(x_1, x_2, x_3) = x_1 = 0,\ e_2(x_1, x_2, x_3) = x_1 - x_2 = 0\}$ be the subspace of $\mathbb{R}^3$.
- $D\mathbf{e}(x_1, x_2, x_3) = \begin{bmatrix} \nabla e_1(x_1, x_2, x_3)^T \\ \nabla e_2(x_1, x_2, x_3)^T \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & -1 & 0 \end{bmatrix}$.
- Because ∇e_1 and ∇e_2 are linearly independent when evaluated at any $(x_1, x_2, x_3) \in S$, all the points of S are regular.
- $T(x_1, x_2, x_3) = \{(y_1, y_2, y_3) \mid \nabla e_1(x_1, x_2, x_3)^T (y_1, y_2, y_3) = 0,\ \nabla e_2(x_1, x_2, x_3)^T (y_1, y_2, y_3) = 0\} = \left\{ (y_1, y_2, y_3) \mid \begin{bmatrix} 1 & 0 & 0 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right\} = \{[0, 0, \alpha] \mid \alpha \in \mathbb{R}\}$=$x_3$ axis in $\mathbb{R}^3$.
- Tangent space at any $\mathbf{x}$ is a one dimensional subspace of $\mathbb{R}^3$.