Optimization Methods (CS1.404), Spring 2024
Lecture 21

Naresh Manwani

Machine Learning Lab, IIIT-H

April 6th, 2024

General Nonlinear Constrained Optimization Problems

Consider the optimization problem

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

where $f(\mathbf{x})$, $h_1(\mathbf{x}), \ldots, h_l(\mathbf{x})$, $e_1(\mathbf{x}), \ldots, e_m(\mathbf{x})$ are smooth functions over $\mathbb{R}^n$.

- Let $\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^n \mid h_j(\mathbf{x}) \leq 0, \ j = 1\ldots l; \ e_i(\mathbf{x}) = 0; \ i = 1\ldots m\}$ be the feasible set.
- Let $\mathbf{x}^* \in \mathcal{X}$ and $\mathcal{A}(\mathbf{x}^*)$ denote set of active constraints at $\mathbf{x}^*$. Then, $\mathcal{A}(\mathbf{x}^*) = I(\mathbf{x}^*) \cup \{1, \ldots, m\}$, where $I(\mathbf{x}^*) = \{j \in \{1, \ldots, l\} \mid h_j(\mathbf{x}^*) = 0\}$.

Regular Point for General Constraint Problems

Definition

A point $\mathbf{x}^* \in \mathcal{X}$ is said to be a regular point, if the gradient vectors $\nabla h_j(\mathbf{x}^*),\ j \in I(\mathbf{x}^*)$ and $\nabla e_i(\mathbf{x}^*),\ i \in \{1, \ldots, m\}$ are linearly independent, where $I(\mathbf{x}^*) = \{j \in \{1, \ldots, l\} \mid h_j(\mathbf{x}^*) = 0\}$. Which means,

$$\sum_{j \in I(\mathbf{x}^*)} \lambda_j \nabla h_j(\mathbf{x}^*) + \sum_{i=1}^{m} \mu_i \nabla e_i(\mathbf{x}^*) = \mathbf{0}$$

only if $\lambda_j = 0,\ j \in I(\mathbf{x}^*)$ and $\mu_i = 0,\ i = 1 \ldots m$.

KKT Necessary Conditions of First Order

Theorem

Consider the optimization problem

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

$$s.t. \;\; h_j(\mathbf{x}) \leq 0, \; j = 1\ldots l$$

$$e_i(\mathbf{x}) = 0; \; i = 1\ldots m$$

where $f(\mathbf{x}), h_1(\mathbf{x}), \ldots, h_l(\mathbf{x}), e_1(\mathbf{x}), \ldots, e_m(\mathbf{x})$ are smooth functions over $\mathbb{R}^n$. If $\mathbf{x}^* \in \mathcal{X}$ is a local minimum and a regular point, then there exist unique vectors $\boldsymbol{\lambda}^* = [\lambda_1^* \; \ldots \; \lambda_l^*]^\top \in \mathbb{R}_+^l$ and $\boldsymbol{\mu}^* = [\mu_1^* \; \ldots \; \mu_m^*]^\top \in \mathbb{R}^m$, such that

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

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

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

$$e_i(\mathbf{x}^*) = 0, \; i = 1\ldots m$$

KKT Point: A point $(\mathbf{x}^* \in \mathcal{X}, \boldsymbol{\lambda} \in \mathbb{R}_+^l, \boldsymbol{\mu} \in \mathbb{R}^m)$ satisfying above conditions is called KKT point.

Proof

- Let $\mathbf{x}^*$ be a regular local minimizer of f on the set $\{\mathbf{x} \in \mathbb{R}^n \mid h_j(\mathbf{x}) \leq 0,\ j = 1 \dots l;\ e_i(\mathbf{x}) = 0,\ i = 1 \dots m\}$. Then, $\mathbf{x}^*$ is also a regular minimizer of f on the set $\hat{S} = \{\mathbf{x} \in \mathbb{R}^n \mid h_j(\mathbf{x}) \leq 0,\ j \in I(\mathbf{x}^*);\ e_i(\mathbf{x}) = 0,\ i = 1 \dots m\}$, where $I(\mathbf{x}^*) = \{j \in \{1, \dots, l\} \mid h_j(\mathbf{x}^*) = 0\}$. Note that the latter set only contains equality constraints. Therefore, from Lagrange's theorem, there exist vectors $\boldsymbol{\lambda}^* \in \mathbb{R}^l$ and $\boldsymbol{\mu}^* \in \mathbb{R}^m$ such that

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

  where for all $j \notin I(\mathbf{x}^*)$, we have $\lambda_j^* = 0$. To complete the proof, we have to show that for all $j \in I(\mathbf{x}^*)$, we have $\lambda_j^* \geq 0$.

- We prove it by contradiction. Suppose that there exists a $p \in I(\mathbf{x}^*)$ such that $\lambda_p^* < 0$.
- Let $\hat{S} = \{\mathbf{x} \in \mathbb{R}^n \mid h_j(\mathbf{x}) = 0,\ j \in I(\mathbf{x}^*),\ j \neq p;\ e_i(\mathbf{x}) = 0,\ i = 1 \dots m\}$ be the surface of active equality constraints except h_p.
- Let $\hat{T}(\mathbf{x}^*) = \{\mathbf{y} \in \mathbb{R}^n \mid \nabla h_j(\mathbf{x})^\top \mathbf{y} = 0,\ j \in I(\mathbf{x}^*),\ j \neq p;\ \nabla e_i(\mathbf{x})^\top \mathbf{y} = 0,\ i = 1 \dots m\}$ be the tangent space corresponding to $\hat{S}$.

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY HYDERABAD

Proof-Continue

- By regularity of $\mathbf{x}^*$, we claim that there exists $\mathbf{y} \in \hat{T}(\mathbf{x}^*)$ such that $\nabla h_p(\mathbf{x}^*)^\top \mathbf{y} \neq 0$.
  - To see this, suppose that for all $\mathbf{y} \in \hat{T}(\mathbf{x}^*)$, $\nabla h_p(\mathbf{x}^*)^\top \mathbf{y} = 0$. This means that, $\nabla h_p(\mathbf{x}^*) \in \hat{T}(\mathbf{x}^*)^\perp$. This implies that $\nabla h_p(\mathbf{x}^*) \in \text{span}[\nabla h_j(\mathbf{x}^*), j \in I(\mathbf{x}^*), j \neq p, \nabla e_i(\mathbf{x}^*), i = 1, \ldots, m]$.
  - But, this contradicts the fact that $\mathbf{x}^*$ is a regular point.
- Without loss of generality, we assume that we have $\mathbf{y}$ such that $\nabla h_p(\mathbf{x}^*)^\top \mathbf{y} < 0$.
- Consider the Lagrange condition, rewritten as

$$\nabla f(\mathbf{x}^*) + \sum_{j \neq p} \lambda_j^* \nabla h_j(\mathbf{x}^*) + \lambda_p^* \nabla h_p(\mathbf{x}^*) + \sum_{i=1}^{m} \mu_i^* \nabla e_i(\mathbf{x}^*) = \mathbf{0}$$

- We take inner product with $\mathbf{y}$ on both sides and get

$$\nabla f(\mathbf{x}^*)^\top \mathbf{y} = -\lambda_p^* \nabla h_p(\mathbf{x}^*)^\top \mathbf{y}$$

- Because, we have assumed that $\nabla h_p(\mathbf{x}^*)^\top \mathbf{y} < 0$ and $\lambda_p^* < 0$, we have $\nabla f(\mathbf{x}^*)^\top \mathbf{y} < 0$.

- Because $\mathbf{y} \in \hat{T}(\mathbf{x}^*)$, we can find a differentiable curve $\{\mathbf{x}(t) \mid t \in (a, b)\}$ on $\hat{S}$ such that there exist $t^* \in (a, b)$ with $\mathbf{x}(t^*) = \mathbf{x}^*$ and $\mathbf{x}'(t^*) = \mathbf{y}$. Now,

$$\frac{d}{dt} f(\mathbf{x}(t^*)) = \nabla f(\mathbf{x}^*)^\top \mathbf{y} < 0.$$

- This means, there exists $\delta > 0$ such that for all $t \in (t^*, t^* + \delta]$, we have

$$f(\mathbf{x}(t)) < f(\mathbf{x}(t^*)) = f(\mathbf{x}^*)$$

- On the other hand, $\frac{d}{dt} h_p(\mathbf{x}(t^*)) = \nabla h_p(\mathbf{x}^*)^\top \mathbf{y} < 0$ and for some $\epsilon > 0$ and all $t \in [t^* + t^* + \epsilon]$, we have that $h_p(\mathbf{x}(t)) \leq 0$.
- Therefore, for all $t \in (t^*, t^* + \min(\delta, \epsilon)]$, we have that $h_p(\mathbf{x}(t)) \leq 0$ and $f(\mathbf{x}(t)) < f(\mathbf{x}(t^*)) = f(\mathbf{x}^*)$.
- Because the points $\mathbf{x}(t)$, $t \in (t^*, t^* + \min(\delta, \epsilon)]$ are in $\hat{S}$, they are feasible points with lower objective function values than $\mathbf{x}^*$.
- This contradicts the assumption that $\mathbf{x}^*$ is a local minimizer, and hence the proof is completed.

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

Theorem

Consider the convex optimization problem

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

where

- $f(\mathbf{x}), h_1(\mathbf{x}), \dots, h_l(\mathbf{x})$ are smooth convex functions over $\mathbb{R}^n$.
- $e_i(\mathbf{x}) = \mathbf{a}_i^\top \mathbf{x} - b_i, \; i = 1 \dots m$

Let $\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^n \mid h_j(\mathbf{x}) \leq 0, \; j = 1 \dots l; \; e_i(\mathbf{x}) = 0; \; i = 1 \dots m\}$ be the feasible set and it satisfies Slater's Condition. Then first order KKT conditions are necessary and sufficient for a global minima of convex optimization problem above.

Interpretation of Lagrange Multipliers

- Consider the problem $\min \; f(\mathbf{x})$ such that $h_j(\mathbf{x}) \leq 0, \; j = 1 \ldots l$.
- Let $\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^n \mid h_j(\mathbf{x}) \leq 0, \; j = 1 \ldots l\}$ be the feasible set.
- Let $\mathbf{x}^*$ be a local minimum and a regular point. Thus, using KKT conditions, $\nabla f(\mathbf{x}^*) + \sum_{j=1}^{l} \lambda_j^* \nabla h_j(\mathbf{x}^*) = \mathbf{0}$ for some $[\lambda_1 \; \lambda_2 \; \ldots \; \lambda_l] \in \mathbb{R}_+^l$.
- Let h_p for some $p \in I(\mathbf{x}^*)$ is perturbed slightly so that $I(\mathbf{x}^*)$ does not change. Given $\epsilon > 0$, the perturbed constraint is as follows:

$$h_p(\mathbf{x}) \leq \epsilon \|\nabla h_p(\mathbf{x}^*)\|.$$

- Consider the new optimization problem as follows.

$$\begin{aligned} \min_{\mathbf{x} \in \mathbb{R}^n} \quad & f(\mathbf{x}) \\ s.t. \quad & h_j(\mathbf{x}) \leq 0, \; j = 1 \ldots l; \; j \neq p \\ & h_p(\mathbf{x}) \leq \epsilon \|\nabla h_p(\mathbf{x}^*)\| \end{aligned}$$

- Let $\mathbf{x}_\epsilon^*$ be a local minima of the new problem.
- Note that we have assumed that $I(\mathbf{x}_\epsilon^*) = I(\mathbf{x}^*)$.

Interpretation of Lagrange Multipliers-Continue

- For the constraint $h_p(\mathbf{x})$, we have $h_p(\mathbf{x}_\epsilon^*) - h_p(\mathbf{x}^*) = \epsilon \|\nabla h_p(\mathbf{x}^*)\|$. Using first order Taylor series approximation at $\mathbf{x}^*$, we get $h_p(\mathbf{x}_\epsilon^*) - h_p(\mathbf{x}^*) \approx \nabla h_p(\mathbf{x}^*)^T(\mathbf{x}_\epsilon^* - \mathbf{x}^*)$. This gives, $\nabla h_p(\mathbf{x}^*)^T(\mathbf{x}_\epsilon^* - \mathbf{x}^*) \approx \epsilon \|\nabla h_p(\mathbf{x}^*)\|$.
- For other constraints h_j, $j \neq p$, we have $h_j(\mathbf{x}_\epsilon^*) - h_j(\mathbf{x}^*) = 0$. Thus, we can get $\nabla h_j(\mathbf{x}^*)^T(\mathbf{x}_\epsilon^* - \mathbf{x}^*) \approx 0,\ j \neq p$.
- Change in the function value:

$$\begin{aligned} f(\mathbf{x}_\epsilon^*) - f(\mathbf{x}^*) &\approx \nabla f(\mathbf{x}^*)^T(\mathbf{x}_\epsilon^* - \mathbf{x}^*) \\ &= -\sum_{j=1}^{l} \lambda_j^* \nabla h_j(\mathbf{x}^*)^T(\mathbf{x}_\epsilon^* - \mathbf{x}^*) \\ &= -\epsilon \lambda_p^* \|\nabla h_p(\mathbf{x}^*)\| \end{aligned}$$

- Dividing by ϵ on both sides and taking limit $\epsilon \to 0$, we get

$$\lim_{\epsilon \to 0} \frac{f(\mathbf{x}_\epsilon^*) - f(\mathbf{x}^*)}{\epsilon} \approx -\lambda_p^* \|\nabla h_p(\mathbf{x}^*)\|$$
$$\Rightarrow \frac{\partial f}{\partial \epsilon} \propto -\lambda_p^*$$

- Thus, λ_p^* captures the rate of change of function f with respect to the perturbation ϵ in the p^{th} constraint $h_p(\mathbf{x})$.

Constraint Classification

Strongly Active Constraint

A constraint is strongly active if it belongs to $\mathcal{A}(\mathbf{x}^*)$ and it has

- strictly positive Lagrange multiplier for inequality constraint $(\lambda_j > 0)$.
- strictly nonzero Lagrange multiplier for equality constraint $(\mu_i \neq 0)$.

Weakly Active Constraint

A constraint is weakly active at if it belongs to $\mathcal{A}(\mathbf{x}^*)$ and it has a zero-valued Lagrange multiplier $(\lambda_j = 0$ or $\mu_i = 0)$.

Inactive Constraint

An inequality constraint is inactive active at if it does not belongs to $\mathcal{A}(\mathbf{x}^*)$. Thus, it has a zero-valued Lagrange multiplier $(\lambda_j = 0)$.

Weakly Active and Inactive Constraints do not participate !

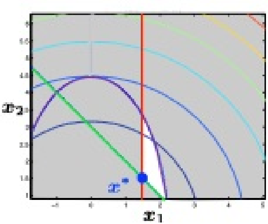
INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY HYDERABAD

Example: Constraint Classification

Consider Optimization Problem as follows.

$$\min_{(x_1,x_2)\in\mathbb{R}^2} \quad x_1^2 + x_2^2$$

$$\begin{aligned} s.t. \quad & h_1(x_1,x_2) = x_1 + x_2 - 3 \geq 0 \qquad \text{(strongly active)} \\ & h_2(x_1,x_2) = x_1 - 1.5 \geq 0 \qquad \text{(weakly active)} \\ & h_3(x_1,x_2) = -x_1^2 - 4x_2^2 + 5 \geq 0 \qquad \text{(inactive)} \end{aligned}$$



The solution is unchanged even if constraints h_2 and h_3 are removed.

Second Order Necessary Conditions

Theorem

Let $\mathbf{x}^*$ be a local minimum of the optimization problem described below.

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

where $f(\mathbf{x}) \in \mathbb{C}^2(\mathbb{R}^n)$, $h_j \in \mathbb{C}^2(\mathbb{R}^n)$, $j = 1 \dots l$ and $e_i \in \mathbb{C}^2(\mathbb{R}^n)$, $i = 1 \dots m$. Suppose that $\mathbf{x}^*$ is regular, which means $\nabla h_j(\mathbf{x}^*)$, $j \in I(\mathbf{x}^*)$ and $\nabla e_i(\mathbf{x}^*)$, $i \in \{1, \dots, m\}$ are linearly independent, where $I(\mathbf{x}^*) = \{j \in \{1, \dots, l\} \mid h_j(\mathbf{x}^*) = 0\}$.

1. Then there exist $\boldsymbol{\lambda}^* = [\lambda_1^* \ \dots \ \lambda_l^*]^\top \in \mathbb{R}^l_+$ and $\boldsymbol{\mu}^* = [\mu_1^* \ \dots \ \mu_m^*]^\top \in \mathbb{R}^m$, such that

$$\begin{aligned}
&\nabla f(\mathbf{x}^*) + \sum_{j=1}^{l} \lambda_j^* \nabla h_j(\mathbf{x}^*) + \sum_{i=1}^{m} \mu_i^* \nabla e_i(\mathbf{x}^*) = \mathbf{0} \\
&\lambda_j^* h_j(\mathbf{x}^*) = 0,\ j = 1 \dots l
\end{aligned}$$

2. and $\mathbf{y}^\top[\nabla^2 f(\mathbf{x}^*) + \sum_{j=1}^{l} \lambda_j^* \nabla h_j^2(\mathbf{x}^*) + \sum_{i=1}^{m} \mu_i^* \nabla e_i^2(\mathbf{x}^*)]\mathbf{y} \geq 0$ for all $\mathbf{y} \in \hat{T}(\mathbf{x}^*)$ where

$$\hat{T}(\mathbf{x}^*) = \{\mathbf{y} \in \mathbb{R}^n \mid \nabla h_j(\mathbf{x}^*)^\top \mathbf{y} = 0,\ j \in I(\mathbf{x}^*);\ \nabla e_i(\mathbf{x}^*)^\top \mathbf{y} = 0,\ i = 1 \dots m\}.$$

Second Order Sufficiency Conditions

Theorem

Consider the optimization problem described below.

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

where $f(\mathbf{x}) \in \mathbb{C}^2(\mathbb{R}^n)$, $h_j \in \mathbb{C}^2(\mathbb{R}^n)$, $j = 1 \dots l$ and $e_i \in \mathbb{C}^2(\mathbb{R}^n)$, $i = 1 \dots m$. Suppose there exist a feasible point $\mathbf{x}^*$, $\boldsymbol{\lambda}^* = [\lambda_1 \; \lambda_2 \; \dots \; \lambda_l]^\top \in \mathbb{R}^l_+$ and $\boldsymbol{\mu} = [\mu_1 \; \mu_2 \; \dots \; \mu_m]^\top \in \mathbb{R}^m$, such that

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

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

2. Also, for all $\mathbf{y} \in \tilde{T}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$, $\mathbf{y} \neq \mathbf{0}$, we have $\mathbf{y}^\top \nabla^2 \mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)\mathbf{y} > 0$. where

$$\tilde{T}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \{\mathbf{y} \in \mathbb{R}^n \mid \nabla h_j(\mathbf{x}^*)^\top \mathbf{y} = 0, \;\; j \in \hat{I}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*); \;\; \nabla e_i(\mathbf{x}^*)^\top \mathbf{y} = 0, \;\; i = 1 \dots m\}.$$

for $\hat{I}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \{j \in \{1, \dots, l\} \mid h_j(\mathbf{x}^*) = 0, \lambda_j^* > 0\}$.

Then $\mathbf{x}^*$ is a local minimizer.

Example 1

- Consider the problem

$$\begin{aligned} \min \ & (x_1 - 1)^2 + x_2 - 2 \\ & h(x_1, x_2) = x_2 - x_1 - 1 = 0 \\ & g(x_1, x_2) = x_1 + x_2 - 2 \leq 0 \end{aligned}$$

- For all $(x_1, x_2) \in \mathbb{R}^2$, we have $\nabla h(x_1, x_2) = [-1 \ \ 1]^\top$ and $\nabla g(x_1, x_2) = [1 \ \ 1]^\top$.
- Thus, $\nabla h(x_1, x_2)$ and $\nabla g(x_1, x_2)$ are linearly independent and hence all feasible points are regular.
- $\nabla f(x_1, x_2) = [2(x_1 - 1) \ \ 1]^\top$.
- KKT conditions are as follows.

$$\begin{aligned} & \nabla f(x_1, x_2) + \lambda \nabla g(x_1, x_2) + \mu \nabla h(x_1, x_2) = [2x_1 - 2 - \mu + \lambda, 1 + \mu + \lambda]^\top = [0, 0]^\top \\ & \lambda(x_1 + x_2 - 2) = 0 \\ & \lambda \geq 0 \\ & x_2 - x_1 - 1 = 0 \\ & x_1 + x_2 - 2 \leq 0 \end{aligned}$$

- To find points that satisfy above conditions, we analyse two cases: (a) $\lambda > 0$, (b) $\lambda = 0$.

Example 1 - Case 1 ($\lambda > 0$)

- $\lambda > 0$ implies that $x_1 + x_2 - 2 = 0$. Thus, we are faced with a system of four linear equations.

$$
\begin{aligned}
&2x_1 - 2 - \mu + \lambda = 0\\
&1 + \mu + \lambda = 0\\
&x_2 - x_1 - 1 = 0\\
&x_1 + x_2 - 2 = 0
\end{aligned}
$$

- Solving the above system of equations, we obtain $x_1 = \frac{1}{2},\ x_2 = \frac{3}{2},\ \lambda = 0,\ \mu = -1$.
- However, this is not a legitimate solution to KKT condition, because we obtain $\lambda = 0$, which contradicts the assumption that $\lambda > 0$.

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY HYDERABAD

Example 1 - Case 2 ($\lambda = 0$)

- Assuming $\lambda = 0$, we are faced with a system of three linear equations.

$$2x_1 - 2 - \mu = 0$$
$$1 + \mu = 0$$
$$x_2 - x_1 - 1 = 0$$

And the solution must satisfy $x_1 + x_2 - 2 \leq 0$.

- Solving the above system of equations, we obtain $x_1 = \frac{1}{2},\ x_2 = \frac{3}{2},\ \mu = -1$.
- Note that $(x_1^*, x_2^*) = [\frac{1}{2}, \frac{3}{2}]^\top$ satisfy the constraint $x_1 + x_2 - 2 \leq 0$.
- $(x_1^*, x_2^*) = [\frac{1}{2}, \frac{3}{2}]^\top$ is a candidate for being a minimizer.
- We now verify that the point $(x_1^*, x_2^*) = [\frac{1}{2}, \frac{3}{2}]^\top$, $\lambda^* = 0$ and $\mu^* = -1$ satisfy the second order sufficient conditions.
- For this, we form the matrix

$$\begin{aligned}\nabla^2 \mathcal{L}(x_1^*, x_2^*, \lambda^*, \mu^*) &= \nabla^2 f(x_1^*, x_2^*) + \lambda^* \nabla^2 h(x_1^*, x_2^*) + \mu^* \nabla g(x_1^*, x_2^*) \\ &= \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}\end{aligned}$$

- We then find the subspace $\tilde{T}(x_1^*, x_2^*, \lambda^*, \mu^*) = \{\mathbf{y} \mid \nabla h(x_1^*, x_2^*)^\top \mathbf{y} = 0\}$.
- Note that $\lambda^* = 0$, the active constraint $x_1 + x_2 = 2$ does not enter into the computation of $\tilde{T}(x_1^*, x_2^*, \lambda^*, \mu^*)$.

Example 1 - Case 2 ($\lambda = 0$)

- We have $\tilde{T}(x_1^*, x_2^*, \lambda^*, \mu^*) = \{\mathbf{y} \mid [-1, 1]\mathbf{y} = 0\} = \{[a, a]^\top \mid a \in \mathbb{R}\}$.
- We then check for positive semi-definiteness of $\nabla^2 \mathcal{L}(x_1^*, x_2^*, \lambda^*, \mu^*)$ on $\hat{T}(x_1^*, x_2^*, \lambda^*, \mu^*)$.
- We have $\mathbf{y}^\top \nabla^2 \mathcal{L}(x_1^*, x_2^*, \lambda^*, \mu^*)\mathbf{y} = [a, a] \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} a \\ a \end{bmatrix} = 2a^2$.
- Thus, $\nabla^2 \mathcal{L}(x_1^*, x_2^*, \lambda^*, \mu^*)$ is positive definite on $\tilde{T}(x_1^*, x_2^*, \lambda^*, \mu^*)$.
- By second order sufficient conditions, we conclude that $(x_1^*, x_2^*) = [\frac{1}{2}, \frac{3}{2}]^\top$ is a strict local minimizer.