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

Naresh Manwani

Machine Learning Lab, IIIT-H

March 28th, 2024

Equality Constraint Problems

Optimization problem with equality constraints is given as below.

$$\min_{\mathbf{x}\in\mathbb{R}^n} \; f(\mathbf{x})$$
$$s.t. \;\; e_i(\mathbf{x}) = 0; \; i = 1 \ldots m$$

where $f(\mathbf{x})$, $e_1(\mathbf{x}), \ldots, e_m(\mathbf{x})$ 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.

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

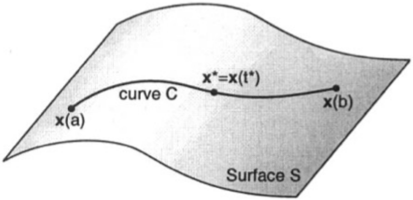
$$S = \{\mathbf{x} \in \mathbb{R}^n \mid e_1(\mathbf{x}^*) = 0, \ldots, e_m(\mathbf{x}^*) = 0\}.$$

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

Curve on the Surface

Definition

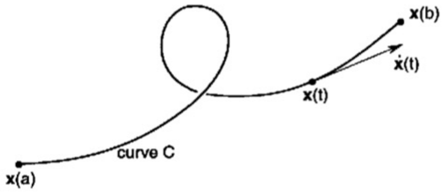
A curve C on a surface S is a set of points $\{\mathbf{x}(t) \in S \mid t \in (a, b)\}$, continuously parameterized by $t \in (a, b)$, that is, $\mathbf{x} : (a, b) \to S$ is a continuous function.



- 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 $\mathbf{x}(t)$.

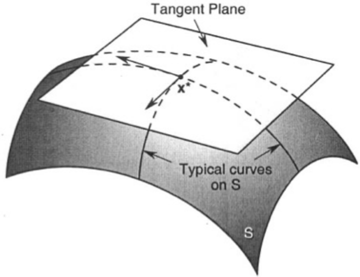


Tangent Space

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.

Key Result: Gradient is Perpendicular to the Level Curve

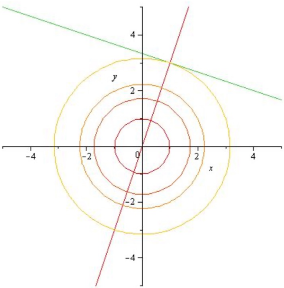
- The derivative of the curve on a surface at a point is a tangent vector to the curve, hence to the surface.
- This intuition agrees with the our definition of Tangent space whenever $\mathbf{x}^*$ is regular.

Theorem

Suppose $\mathbf{x}^* \in S$ is a regular point, and $T(\mathbf{x}^*)$ is a tangent space at $\mathbf{x}^*$. Then, $\mathbf{y} \in T(\mathbf{x}^*)$ if and only if there exists a differentiable curve in S passing through $\mathbf{x}^*$ with derivative $\mathbf{y}$ at $\mathbf{x}^*$.

Example

- Let $e(x,y) = x^2 + y^2$. $\nabla e(x,y) = \begin{pmatrix} 2x \\ 2y \end{pmatrix}$.
- Consider level curve $e(x,y) = 10$. Consider a point $(1,3)$ on it. $\nabla e(1,3) = \begin{pmatrix} 2 \\ 6 \end{pmatrix}$.
- To find the slope of the tangent for the level curve, we differentiate it with respect to x. Which gives, $2x + 2y\frac{\partial y}{\partial x} = 0$ or $\frac{\partial y}{\partial x} = -\frac{x}{y}$.
- At (1,3), the slope of the tangent curve is $\frac{\partial y}{\partial x} = -\frac{1}{3}$. In the vector form, it is $\mathbf{z} = \begin{pmatrix} 3 \\ -1 \end{pmatrix}$.
- It is easy to see that $\mathbf{z}^T \nabla e(1,3) = 0$.



Normal Space

Definition

The normal space $N(\mathbf{x}^*)$ 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

$$N(\mathbf{x}^*) = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} = D\mathbf{e}(\mathbf{x}^*)\mathbf{z},\ \mathbf{z} \in \mathbb{R}^m\}$$

- Thus, $N(\mathbf{x}^*)$ is the range of $D\mathbf{e}(\mathbf{x}^*)^T$.
- Thus, normal space is the subspace of $\mathbb{R}^n$ spanned by the vectors $\nabla e_1(\mathbf{x}^*), \ldots, \nabla e_m(\mathbf{x}^*)$.
- Normal space $N(\mathbf{x}^*)$ contains origin into it.
- Assuming $\mathbf{x}^*$ a regular point, dimension of the normal space $N(\mathbf{x}^*)$ is m.

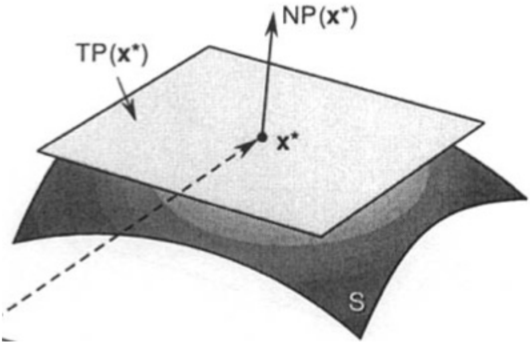
Normal Plane

Normal Plane

We define the **Normal Plane** at $\mathbf{x}^*$ as the set

$$NP(\mathbf{x}^*) = N(\mathbf{x}^*) + \mathbf{x}^* = \{\mathbf{x} + \mathbf{x}^* \in \mathbb{R}^n \mid \mathbf{x} \in N(\mathbf{x}^*)\}$$

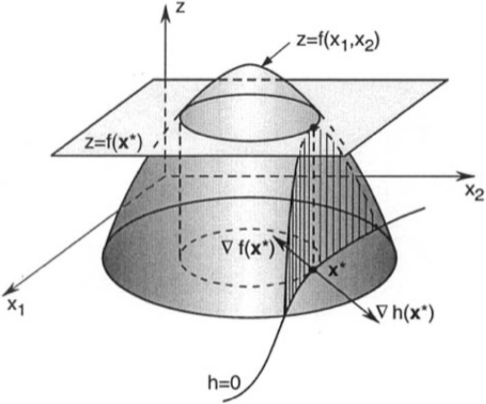
We have, $T(\mathbf{x}^*) = N(\mathbf{x}^*)^\top$ and $N(\mathbf{x}^*) = T(\mathbf{x}^*)^\top$.



Lagrange's Optimality Condition for $n = 2$ and $m = 1$

Theorem

Let $\mathbf{x}^*$ be a local minimizer of $f : \mathbb{R}^2 \to \mathbb{R}$ subject to the constraint $e(\mathbf{x}) = 0$ where $e : \mathbb{R}^2 \to \mathbb{R}$. Then $\nabla f(\mathbf{x}^*)$ and $\nabla e(\mathbf{x}^*)$ are parallel. That is, if $\nabla e(\mathbf{x}^*) \neq \mathbf{0}$, then there exist a scalar μ^* such that $\nabla f(\mathbf{x}^*) + \mu^* \nabla e(\mathbf{x}^*) = \mathbf{0}$. Here, we refer to μ^* as the Lagrange multiplier.



Lagrange's Conditions are Only Necessary, not Sufficient

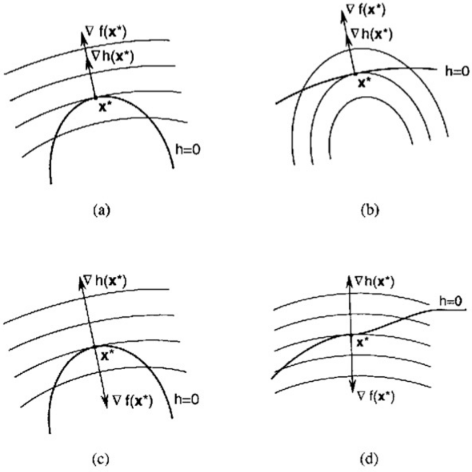


Figure: (a) maximizer, (b) minimizer, (c) minimizer, (d) not an extremizer.

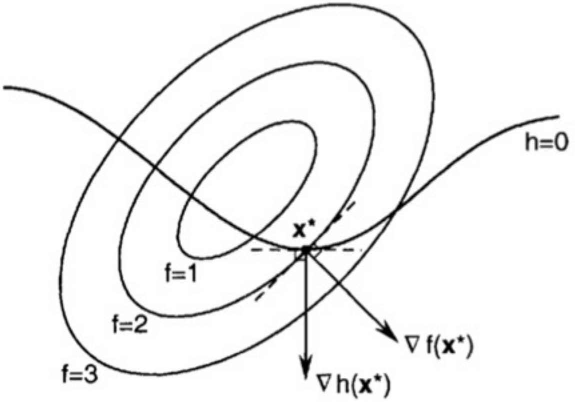
Lagrange's Theorem: General Case

Theorem

Let $\mathbf{x}^*$ be a local minimizer of $f : \mathbb{R}^n \to \mathbb{R}$ subject to the constraint $\mathbf{e}(\mathbf{x}) = 0$ where $\mathbf{e} : \mathbb{R}^n \to \mathbb{R}^m$ $(m \leq n)$. Assume that $\mathbf{x}^*$ is a regular point. Then there exist a scalar $\boldsymbol{\mu}^* \in \mathbb{R}^m$ such that $\nabla f(\mathbf{x}^*) + \boldsymbol{\mu}^* D\mathbf{e}(\mathbf{x}^*) = \mathbf{0}^T$. Here, we refer to $\boldsymbol{\mu}^*$ as the Lagrange multiplier vector.

When Lagrange Condition Does Not Hold?

$\mathbf{x}^*$ can not be an extremizer if $\nabla f(\mathbf{x}^*) \notin N(\mathbf{x}^*)$.



Lagrangian Function

- Lagrangian function $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is given by:

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\mu}) = f(\mathbf{x}) + \boldsymbol{\mu}^T \mathbf{e}(\mathbf{x})$$

- The Lagrange condition for a local minimizer $\mathbf{x}^*$ can be represented as

$$D\mathcal{L}(\mathbf{x}^*, \boldsymbol{\mu}^*) = \mathbf{0}^T$$

for some $\boldsymbol{\mu}^* \in \mathbb{R}^m$, where the derivative operation D is with respect to the entire argument $(\mathbf{x}, \boldsymbol{\mu})$.

- Which is equivalent to solving the equations

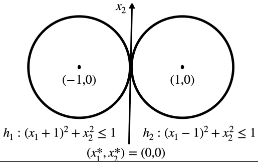
$$D_{\mathbf{x}}\mathcal{L}(\mathbf{x}, \boldsymbol{\mu}) = \mathbf{0}^T$$
$$D_{\boldsymbol{\mu}}\mathcal{L}(\mathbf{x}, \boldsymbol{\mu}) = \mathbf{0}^T$$

- The above represents $n + m$ equations in $n + m$ unknowns.
- Note that Lagrange condition is only necessary.

- Consider the optimization problem as follows.

$$\min \ x_1 - 3x_2$$
$$e_1(\mathbf{x}) = (x_1 - 1)^2 + x_2^2 = 1$$
$$e_2(\mathbf{x}) = (x_1 + 1)^2 + x_2^2 = 1$$

- Here, Feasible set $\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^2 \mid (x_1 + 1)^2 + x_2^2 = 1,\ (x_1 - 1)^2 + x_2^2 = 1\} = \{(0,0)\}$.
- $\mathcal{L}(\mathbf{x}, \mu_1, \mu_2) = x_1 - 3x_2 + \mu_1[(x_1 - 1)^2 + x_2^2 - 1] + \mu_2[(x_1 + 1)^2 + x_2^2 - 1]$.
- $\nabla f(\mathbf{x}) = \begin{bmatrix} 1 \\ -3 \end{bmatrix}$, $\nabla e_1(\mathbf{x}) = \begin{bmatrix} 2(x_1 - 1) \\ 2x_2 \end{bmatrix}$, $\nabla e_2(\mathbf{x}) = \begin{bmatrix} 2(x_1 + 1) \\ 2x_2 \end{bmatrix}$.
- $\nabla f(0,0) = \begin{bmatrix} 1 \\ -3 \end{bmatrix}$, $\nabla e_1(0,0) = \begin{bmatrix} -2 \\ 0 \end{bmatrix}$, $\nabla e_2(\mathbf{x}) = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$.
- If $\mathbf{x}^* = (0,0)$ is a local minima, then $\nabla\mathcal{L}(\mathbf{x}^*, \mu_1, \mu_2) = \mathbf{0}$. Which implies, $1 - 2\mu_1 + 2\mu_2 = 0$ and $-3 = 0$, which is impossible.
- Thus, $\mathbf{x}^* = (0,0)$ is not a local minima.
- Note that $\mathbf{x}^* = (0,0)$ is not a regular point.



Example 2

- Consider the following problem: $\max_{\mathbf{x}\in\mathbb{R}^n} \frac{\mathbf{x}^T Q\mathbf{x}}{\mathbf{x}^T\mathbf{x}}$ where Q is a symmetric positive semi-definite matrix.
- Note that is $\mathbf{x}$ is a solution to the problem, then $t\mathbf{x}$ is also a solution for any $t \neq 0$. $\left(\frac{(t\mathbf{x})^T Q(t\mathbf{x})}{(t\mathbf{x})^T(t\mathbf{x})} = \frac{\mathbf{x}^T Q\mathbf{x}}{\mathbf{x}^T\mathbf{x}}\right)$.
- To avoid the multiplicity of the solutions, we add the constraint $\mathbf{x}^T\mathbf{x} = 1$.
- Thus,

$$\max_{\mathbf{x}\in\mathbb{R}^n} \frac{\mathbf{x}^T Q\mathbf{x}}{\mathbf{x}^T\mathbf{x}} = \max_{\mathbf{x}\in\mathbb{R}^n} \ \mathbf{x}^T Q\mathbf{x}$$

$$s.t.\ \mathbf{x}^T\mathbf{x} = 1$$

- So, $f(\mathbf{x}) = \mathbf{x}^T Q\mathbf{x}$ and $e(\mathbf{x}) = 1 - \mathbf{x}^T\mathbf{x}$.
- Any feasible point for this problem is regular.
- Lagrange conditions yield $2\mathbf{x}^T Q - 2\mu\mathbf{x}^T = 0$ and $1 - \mathbf{x}^T\mathbf{x} = 0$.
- The first condition gives $Q\mathbf{x} = \mu\mathbf{x}$. Therefore, the solution, if exists, is an eigen vector of Q.
- Let $\mathbf{x}^*$ and μ^* be the optimal solution. Because $(\mathbf{x}^*)^T\mathbf{x}^* = 1$ and $Q\mathbf{x}^* = \mu^*\mathbf{x}^*$.. This gives

$$\mu^* = (\mathbf{x}^*)^T Q\mathbf{x}^*$$

- Hence μ^* is the maximum of the objective function, and therefore, the maximum eigen value of Q.