Applied Linear Algebra in Data Analysis

Introduction to Constrained Optimization

Sivakumar Balasubramanian

Department of Bioengineering
Christian Medical College, Bagayam
Vellore 632002

Constrained Optimization

- A general optimization problem can be fomulated as the following,

$$
\begin{aligned}
\text{minimize} \quad & f(\mathbf{x}) \\
\text{subject to} \quad & \mathbf{g}(\mathbf{x}) \leq \mathbf{0}, \;\; \mathbf{g}(\mathbf{x}) = \begin{bmatrix} g_1(\mathbf{x}) & g_2(\mathbf{x}) & \cdots & g_p(\mathbf{x}) \end{bmatrix}^\top \\
& \mathbf{h}(\mathbf{x}) = \mathbf{0}, \;\; \mathbf{h}(\mathbf{x}) = \begin{bmatrix} h_1(\mathbf{x}) & h_2(\mathbf{x}) & \cdots & h_q(\mathbf{x}) \end{bmatrix}^\top
\end{aligned}
$$

where, $f(\mathbf{x})$ is the **objective function** and $\mathbf{g}(\mathbf{x})$ represents the set of **inquality constaints** and $\mathbf{h}(\mathbf{x})$ represents the set of **equality constraints**.

Constrained Optimization

The set of all values of $\mathbf{x}$ that satisfy the constraints is called the **feasible set** and is denoted by $\mathcal{F}$.

$$\mathcal{F} = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{g}(\mathbf{x}) \leq \mathbf{0} \text{ and } \mathbf{h}(\mathbf{x}) = \mathbf{0}\}$$

Whatever the form, with constrained optimization the search for the minimizer of the function $f(\mathbf{x})$ is restricted to the feasible set $\mathcal{F}$.

Feasible direction - A direction $\mathbf{d}$ is said to be a feasible direction at the point $\mathbf{x} \in \mathcal{F}$ if there exists $\alpha_0 > 0$ such that $\mathbf{x} + \alpha\mathbf{d} \in \mathcal{F}$ for all $\alpha \in [0, \alpha_0]$.

The conditions for a point to be a minimizer in the constrained optimization case are different from the unconstrained case.

First Order Necessary Condition. Let $\mathcal{F} \in \mathbb{R}^n$, and let $f : \mathcal{F} \to \mathbb{R}$ be a differentiable function (at least once). If $\mathbf{x}^\star$ is a local minimizer of the function f over the set $\mathcal{F}$, then for any feasible direction $\mathbf{d}$ at the point $\mathbf{x}^\star$, we have,

$$\mathbf{d}^\top \nabla f\left(\mathbf{x}^\star\right) \geq 0$$

This above conditions applies for both points at the boundary and interior to the set $\mathcal{F}$.

For interior points, all directions are feasible, and so the above condition becomes,

$$\nabla f\left(\mathbf{x}^\star\right) = 0$$

which is the same as the unconstrained case.

Constrained Optimization

Second Order Necessary Condition. Let $\mathcal{F} \in \mathbb{R}^n$, and let $f : \mathcal{F} \to \mathbb{R}$ be a differentiable function (at least twice), $\mathbf{x}^\star$ is a local minimizer of f over $\mathcal{F}$, and $\mathbf{d}$ is a feasible direction at $\mathbf{x}^\star$. If $\mathbf{d}^\top \nabla f(\mathbf{x}) = 0$, then

$$\mathbf{d}^\top \mathbf{H}(\mathbf{x}^\star)\,\mathbf{d} \geq 0$$

where $\mathbf{H}(\mathbf{x})$ is the Hessian of f at $\mathbf{x}$.

Constrained Optimization: Equality Constraints

Consider the following optimization problem with only equality constraints,

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{h}(\mathbf{x}) = \mathbf{0} \end{aligned}$$

where, $\mathbf{h}(\mathbf{x}) = \begin{bmatrix} h_1(\mathbf{x}) & h_2(\mathbf{x}) & \cdots & h_q(\mathbf{x}) \end{bmatrix}^\top$, and we assume that each one of these constraint functios is differentiable.

The feasible set for this problem is,

$$\mathcal{F} = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{h}(\mathbf{x}) = \mathbf{0}\}$$

Note that the feasiblity set forms a $n - q$ dimensional surface oer manifold in $\mathbb{R}^n$.

Constrained Optimization: Equality Constraints

Regular point – A point $\mathbf{x}^\star \in \mathcal{F}$ is said to be a regular point if the gradients of the equality constraints are linearly independent at the point $\mathbf{x}^\star$.

Note that the gradient of the equality constraints is a matrix of size $q \times n$.

$$\nabla_{\mathbf{x}} \mathbf{h}(\mathbf{x}) = \begin{bmatrix} \nabla_{\mathbf{x}} h_1(\mathbf{x}) \\ \nabla_{\mathbf{x}} h_2(\mathbf{x}) \\ \vdots \\ \nabla_{\mathbf{x}} h_q(\mathbf{x}) \end{bmatrix} \in \mathbb{R}^{q \times n}$$

The gradient $\nabla \mathbf{h}(\mathbf{x})$ has full rank at a regular point.

Constrained Optimization: Equality Constraints

The **tangent space** of a surface is the higher dimensional equivalent of a tagent line to a one dimensional curve.

The **tangent space** at a point $\mathbf{x}^\star \in \mathcal{F}$ is defined as the following set,

$$T\left(\mathbf{x}^\star\right) = \left\{\mathbf{y} \in \mathbb{R}^n \mid \nabla \mathbf{h}\left(\mathbf{x}\right) \mathbf{y} = \mathbf{0}\right\} = \mathcal{N}\left(\nabla \mathbf{h}\left(\mathbf{x}\right)\right)$$

The **tangent plane** at the point $\mathbf{x}^\star$ on the surface $\mathcal{F}$ is defined as,

$$TP\left(\mathbf{x}^\star\right) = T\left(\mathbf{x}^\star\right) + \mathbf{x}^\star = \left\{\mathbf{x} + \mathbf{x}^\star \mid \mathbf{x} \in T\left(\mathbf{x}^\star\right)\right\}$$

Constrained Optimization: Equality Constraints

The **normal space** of a surface is the orthogonal complement of the tangent space.

The **normal space** at a point $\mathbf{x}^\star \in \mathcal{F}$ is defined as the following set,

$$N\left(\mathbf{x}^\star\right) = \left\{\nabla \mathbf{h}\left(\mathbf{x}\right)^\top \mathbf{y} \mid \mathbf{y} \in \mathbb{R}^q\right\} = \mathcal{R}\left(\nabla \mathbf{h}\left(\mathbf{x}\right)^\top\right)$$

The **normal plane** at the point $\mathbf{x}^\star$ on the surface $\mathcal{F}$ is defined as,

$$NP\left(\mathbf{x}^\star\right) = N\left(\mathbf{x}^\star\right) + \mathbf{x}^\star = \left\{\mathbf{x} + \mathbf{x}^\star \mid \mathbf{x} \in N\left(\mathbf{x}^\star\right)\right\}$$

Constrained Optimization: Equality Constraints

Lagrange theorem: Let $\mathbf{x}^\star$ be a local minimizer (or maximizer) of the function $f : \mathbb{R}^n \to \mathbb{R}$, subject to $\mathbf{h}(\mathbf{x}) = \mathbf{0}$, $\mathbf{h} : \mathbb{R}^n \to \mathbb{R}^q$, $q \leq n$. Assume that $\mathbf{x}^\star$ is a regular point of the equality constraints. Then there exists a vector $\boldsymbol{\lambda}^\star \in \mathbb{R}^q$ such that,

$$\nabla f(\mathbf{x}^\star) + \nabla \mathbf{h}(\mathbf{x}^\star)^\top \boldsymbol{\lambda}^\star = \mathbf{0}$$

where, $\boldsymbol{\lambda}^\star$ is called the **Lagrange multiplier vector**.

This theorem states that extremum of this constrained optimization problem occurs when the gradient of f is in the normal space of the surface $\mathcal{F}$, i.e. $\nabla f(\mathbf{x})$ is a linear combination of the gradients of the individual equality constraints $\nabla h_i(\mathbf{x})$, $1 \leq i \leq q$.

This is also equivalent to saying $\nabla f(\mathbf{x}) \in T(\mathbf{x})^\perp$ or $\nabla f(\mathbf{x}) \in N(\mathbf{x})$

Constrained Optimization: Equality Constraints

It is convenient to introduce the **Lagrange function**: $l : \mathbb{R}^n \times \mathbb{R}^q \to \mathbb{R}$,

$$l\left(\mathbf{x}, \boldsymbol{\lambda}\right) = f\left(\mathbf{x}\right) + \boldsymbol{\lambda}^\top \mathbf{f}\left(\mathbf{x}\right)$$

The Lagrange condition for $\mathbf{x}^\star$ to be a local minimizer is the following,

$$\nabla l\left(\mathbf{x}^\star, \boldsymbol{\lambda}^\star\right) = \mathbf{0}$$

for some $\boldsymbol{\lambda}^\star$, where the gradient is with respect to the combined argument $\begin{bmatrix}\mathbf{x}^\top & \boldsymbol{\lambda}^\top\end{bmatrix}$. This is equivalent to the first order necessary condition for the unconstrained optimization problem with respect to $\begin{bmatrix}\mathbf{x}^\top & \boldsymbol{\lambda}^\top\end{bmatrix}$. This condition is necessary but not sufficient.

Constrained Optimization: Equality Constraints

Let f and $\mathbf{h}$ be both twice continuously differentiable,

$$\begin{aligned} l(\mathbf{x}, \boldsymbol{\lambda}) &= f(\mathbf{x}) + \boldsymbol{\lambda}^\top \mathbf{f}(\mathbf{x}) \\ &= f(\mathbf{x}) + \lambda_1 h_1(\mathbf{x}) + \lambda_2 h_2(\mathbf{x}) + \cdots + \lambda_q h_q(\mathbf{x}) \end{aligned}$$

Taking the Hessian on both sides,

$$\mathbf{H}_l(\mathbf{x}) = \mathbf{H}_f(\mathbf{x}) + \sum_{i=1}^{q} \lambda_i \mathbf{H}_{h_i}(\mathbf{x})$$

where $\mathbf{H}_l$ is the Hessian of $l(\mathbf{x}, \boldsymbol{\lambda})$ with respect to $\mathbf{x}$, and $\mathbf{H}_{h_i}$ is the Hessian of the function $h_i(\mathbf{x})$ with respect to $\mathbf{x}$.

Constrained Optimization: Equality Constraints

Second Order Necessary Condition. Let $\mathbf{x}^\star$ be a local minimizer of f subject to the equality constraints $\mathbf{h}(\mathbf{x}) = \mathbf{0}$, $\mathbf{h} : \mathbb{R}^n \to \mathbb{R}^q$, $q \leq n$, and $f, \mathbf{h}$ are twice differentiable, and $\mathbf{x}^\star$ is a regular point. Then, there exists a vector $\boldsymbol{\lambda}^\star \in \mathbb{R}^q$ such that,

1. $\nabla f(\mathbf{x}^\star) + \boldsymbol{\lambda}^{\star\top} \nabla \mathbf{h}(\mathbf{x}) = \mathbf{0}$
2. For all $\mathbf{y} \in T(\mathbf{x}^\star)$, we have $\mathbf{y}^\top \mathbf{L}(\mathbf{x}^\star, \boldsymbol{\lambda}^\star)\, \mathbf{y} \geq 0$.

Second Order Sufficient Condition. $\mathbf{x}^\star$ is strict local minimizer of f subject to the equality constraints $\mathbf{h}(\mathbf{x}) = \mathbf{0}$, $\mathbf{h} : \mathbb{R}^n \to \mathbb{R}^q$,

1. $\nabla f(\mathbf{x}^\star) + \boldsymbol{\lambda}^{\star\top} \nabla \mathbf{h}(\mathbf{x}) = \mathbf{0}$
2. For all $\mathbf{y} \in T(\mathbf{x}^\star)$, we have $\mathbf{y}^\top \mathbf{L}(\mathbf{x}^\star, \boldsymbol{\lambda}^\star)\, \mathbf{y} > 0$.

Constrained Optimization: Inequality Constraints

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{g}(\mathbf{x}) \leq \mathbf{0}, \ \ \mathbf{g}(\mathbf{x}) = \begin{bmatrix} g_1(\mathbf{x}) & g_2(\mathbf{x}) & \cdots & g_p(\mathbf{x}) \end{bmatrix}^\top \\ & \mathbf{h}(\mathbf{x}) = \mathbf{0}, \ \ \mathbf{h}(\mathbf{x}) = \begin{bmatrix} h_1(\mathbf{x}) & h_2(\mathbf{x}) & \cdots & h_q(\mathbf{x}) \end{bmatrix}^\top \end{aligned}$$

where, $f(\mathbf{x})$ is the **objective function** and $\mathbf{g}(\mathbf{x})$ represents the set of **inquality constaints** and $\mathbf{h}(\mathbf{x})$ represents the set of **equality constraints**.

An inequality constraint $g_i(\mathbf{x}) \leq 0$ is said to be **active** at a point $\mathbf{x}^\star$ if $g_i(\mathbf{x}^\star) = 0$. It is **inactive** at $\mathbf{x}^\star$ if $g_i(\mathbf{x}^\star) < 0$.

Constrained Optimization: Inequality Constraints

Regular point Let the point $\mathbf{x}^\star$ satisfy $\mathbf{h}\left(\mathbf{x}^\star\right) = \mathbf{0}$, $\mathbf{g}\left(\mathbf{x}^\star\right) \leq \mathbf{0}$, and let $J\left(\mathbf{x}^\star\right)$ be the index set of active inquality constraints,

$$J\left(\mathbf{x}^\star\right) = \{j \mid g_j\left(\mathbf{x}^\star\right) = 0\}$$

We say the point $\mathbf{x}^\star$ is a **regular point** if the following vectors form a linearly independent set.

$$\nabla h_i\left(\mathbf{x}^\star\right), \nabla g_j\left(\mathbf{x}^\star\right), \; 1 \leq i \leq q, \, j \in J\left(\mathbf{x}^\star\right)$$

Constrained Optimization: Inequality Constraints

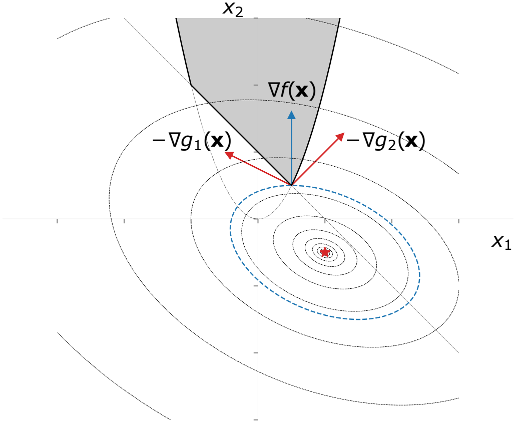
Karush-Kuhn-Tucker Theorem. Let $f, \mathbf{h}, \mathbf{g}$ be differentiable at least once, and let $\mathbf{x}^\star$ be a regular point and a local minimizer of the problem. Then, there exists $\boldsymbol{\lambda}^\star \in \mathbb{R}^q$ and $\boldsymbol{\mu}^\star \in \mathbb{R}^p$ such that,

1. $\boldsymbol{\mu}^\star \geq 0$
2. $\nabla f\left(\mathbf{x}^\star\right) + \nabla \mathbf{h}\left(\mathbf{x}^\star\right)^\top \boldsymbol{\lambda}^\star + \nabla \mathbf{g}\left(\mathbf{x}^\star\right)^\top \boldsymbol{\mu}^\star = \mathbf{0}$
3. $\boldsymbol{\mu}^{\star\top} \mathbf{g}\left(\mathbf{x}^\star\right) = 0$

The vector $\boldsymbol{\lambda}^\star$ is called the **Lagrange multiplier vector** and $\boldsymbol{\mu}^\star$ is called the **Karush-Kuhn-Tucker (KKT) multiplier vector**.

Constrained Optimization: Inequality Constraints

$$f(\mathbf{x}) = \mathbf{x}^\top \mathbf{A}\mathbf{x}$$
$$g_1(\mathbf{x}) = -x_1^2 + x_2 \leq 0$$
$$g_2(\mathbf{x}) = -x_1 + x_2 - 2 \leq 0$$



Constrained Optimization: Inequality Constraints

Tangent space: The tangent space at a point $\mathbf{x}^\star$ is defined as,

$$T\left(\mathbf{x}^\star\right) = \left\{\mathbf{y} \in \mathbb{R}^n \mid \nabla \mathbf{h}\left(\mathbf{x}^\star\right)\mathbf{y} = \mathbf{0}, \nabla g_j\left(\mathbf{x}^\star\right)^\top \mathbf{y} = 0,\ j \in J\left(\mathbf{x}^\star\right)\right\}$$

Second order necessary condition. Let $\mathbf{x}^\star$ be a local minimizer of $f\left(\mathbf{x}\right)$ subject to $\mathbf{h}\left(\mathbf{x}^\star\right) = \mathbf{0}$ and $\mathbf{g}\left(\mathbf{x}\right) \leq \mathbf{0}$. Let $f, \mathbf{h}, \mathbf{g}$ be at least twice differentiableThen, there exists $\boldsymbol{\lambda}^\star \in \mathbb{R}^q$ and $\boldsymbol{\mu}^\star \in \mathbb{R}^p$ such that,

1. $\nabla f\left(\mathbf{x}^\star\right) + \nabla \mathbf{h}\left(\mathbf{x}^\star\right)^\top \boldsymbol{\lambda}^\star + \nabla \mathbf{g}\left(\mathbf{x}^\star\right)^\top \boldsymbol{\mu}^\star = \mathbf{0}$
2. For all $\mathbf{y} \in T\left(\mathbf{x}^\star\right)$, we have $\mathbf{y}^\top \mathbf{H}\left(\mathbf{x}^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\mu}^\star\right)\mathbf{y} \geq 0$.

where,

$$\mathbf{H}\left(\mathbf{x}\right) = \mathbf{H}_f\left(\mathbf{x}\right) + \sum_{i=1}^{p} \mu_i \mathbf{H}_{g_i}\left(\mathbf{x}\right) + \sum_{i=1}^{q} \lambda_i \mathbf{H}_{h_i}\left(\mathbf{x}\right)$$

Constrained Optimization: Inequality Constraints

Second order sufficient condition. Let $\mathbf{x}^\star$ be a local minimizer of $f(\mathbf{x})$ subject to $\mathbf{h}(\mathbf{x}^\star) = \mathbf{0}$ and $\mathbf{g}(\mathbf{x}) \leq \mathbf{0}$. Let $f, \mathbf{h}, \mathbf{g}$ be at least twice differentiableThen, there exists $\boldsymbol{\lambda}^\star \in \mathbb{R}^q$ and $\boldsymbol{\mu}^\star \in \mathbb{R}^p$ such that,

1. $\nabla f(\mathbf{x}^\star) + \nabla \mathbf{h}(\mathbf{x}^\star)^\top \boldsymbol{\lambda}^\star + \nabla \mathbf{g}(\mathbf{x}^\star)^\top \boldsymbol{\mu}^\star = \mathbf{0}$
2. For all $\mathbf{y} \in \tilde{T}(\mathbf{x}^\star)$, we have $\mathbf{y}^\top \mathbf{H}(\mathbf{x}^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\mu}^\star)\,\mathbf{y} \geq 0$.

where,

$$\mathbf{H}(\mathbf{x}) = \mathbf{H}_f(\mathbf{x}) + \sum_{i=1}^{p} \mu_i \mathbf{H}_{g_i}(\mathbf{x}) + \sum_{i=1}^{q} \lambda_i \mathbf{H}_{h_i}(\mathbf{x})$$

and,

$$\tilde{T}(\mathbf{x}^\star) = \{\}$$

$\mathbf{x}^\star$ is a strict local minimizer of $f(\mathbf{x})$ subject to $\mathbf{h}(x) = \mathbf{0}$ and $\mathbf{g}(\mathbf{x}) \leq \mathbf{0}$.