Non-Linear Programming

Constrained optimization: Equality and Inequality type constraints

Anand Jacob Abraham

Department of Industrial & Systems Engineering
IIT Kharagpur

Table of contents

2

Math review: Subspaces associated with a matrix

Null space

Associated with every matrix there are two important subspaces. They are called the Range space/column space and the Null space. They are defined as below.

Definition (Null space)

The null space of a matrix $V \in \mathbb{R}^{m \times n}$ is a set of vectors $v \in \mathbb{R}^n$ such that $Vv = 0$. i.e.,

$$\mathcal{N}(V) = \{v : Vv = 0\}$$

- Multiplication of a vector with a matrix is also called as a *Linear transformation*. i.e, the vector v is transformed from the space $\mathbb{R}^n$ to some other vector space R^m.
- The Null space $\mathcal{N}(V)$ is the set of vectors whose linear transformation yields a zero vector.

Range space

Definition (Range space/Column space)

The range space of a matrix $\boldsymbol{V} \in \mathbb{R}^{m \times n}$ is a set of vectors $\boldsymbol{y} \in \mathbb{R}^m$ such that $\boldsymbol{Vv} = \boldsymbol{y}$. i.e.,

$$\mathcal{R}(\boldsymbol{V}) = \{\boldsymbol{y} : \boldsymbol{Av} = \boldsymbol{y}\}$$

- This is simply the span of the columns of $\boldsymbol{V}$.
- The $dim(\mathcal{R}(\boldsymbol{V})) = rank(\boldsymbol{V})$.
- One can also define a row space like this, it is denoted by $\mathcal{R}(\boldsymbol{V}^T)$.

Definition (Rank of a matrix)

The number linearly independent rows of a matrix is called its row rank. The number of linearly independent columns of the matrix is called its column rank.

Row rank and column rank are numerically always equal and are both generally called *rank*.

Orthogonal complements

Definition (Orthogonal complement)

Given a subspace S, $S^\perp$ denotes a set known as its orthogonal complement if $S^\perp \equiv \{\boldsymbol{v} : \boldsymbol{v}^T \boldsymbol{w} = 0, \forall \boldsymbol{w} \in S\}$. That is, every member of the subspace is perpendicular (orthogonal) to every member of the orthogonal complement subspace.

That is, every member of the subspace is perpendicular (orthogonal) to every member of the orthogonal complement subspace. An additional property related to the orthogonal complement is known as *orthogonal decomposition*. $\forall \boldsymbol{u} \in \mathbb{R}^n$, there exist some $\boldsymbol{v} \in S$ and $\boldsymbol{w} \in S^\perp$ such that,

$$\boldsymbol{u} = \boldsymbol{v} + \boldsymbol{w}.$$

Orthogonal complemantarity of Range and Null Spaces

- For $V \in \mathbb{R}^{m \times n}$, $\mathcal{R}(V) = \mathcal{N}(V^T)^\perp$ and $\mathcal{R}(V^T) = \mathcal{N}(V)^\perp$.

- For $V \in \mathbb{R}^{m \times n}$, always $rank(V) = dim(\mathcal{R}(V)) = dim(\mathcal{R}(V^T))$.

- Suppose for $V \in \mathbb{R}^{m \times n}$, let $rank(V) = r$, then $dim(\mathcal{R}(V)) = dim(\mathcal{R}(V^T)) = r$ and $dim(\mathcal{R}(V)) + dim((\mathcal{N}(V^T)) = m$ (each column $v_i \in \mathbb{R}^m$ as it is an $m \times n$ matrix) and $dim(\mathcal{R}(V^T)) + dim(\mathcal{N}(V)) = n$.

Rank- Nullity Theorem

The $dim(\mathcal{N}(\boldsymbol{V}))$ is called *Nullity* of $\boldsymbol{V}$.

" The rank of a matrix and the dimension of the null space of the matrix (also known as *nullity*) always yields a sum equal to the number of columns".

That is, $rank(\boldsymbol{V}) + dim(\mathcal{N}(\boldsymbol{V})) = n$. This is known as the *Rank-Nullity Theorem* in Linear Algebra.

Regularity

Regular point

Definition (Regular point)

$\boldsymbol{x}^*$ is called a regular point if $\nabla h_1(\boldsymbol{x}^*), \nabla h_2(\boldsymbol{x}^*), \ldots, \nabla h_m(\boldsymbol{x}^*)$ are linearly independent or if the Jacobian $D\boldsymbol{h}(\boldsymbol{x}^*))$ is full row rank.

The Jacobian has the following structure.

$$D\boldsymbol{h}(\boldsymbol{x}^*) = \begin{bmatrix} \nabla^T h_1(\boldsymbol{x}) \\ \nabla^T h_2(\boldsymbol{x}) \\ \vdots \\ \nabla^T h_m(\boldsymbol{x}) \end{bmatrix}$$
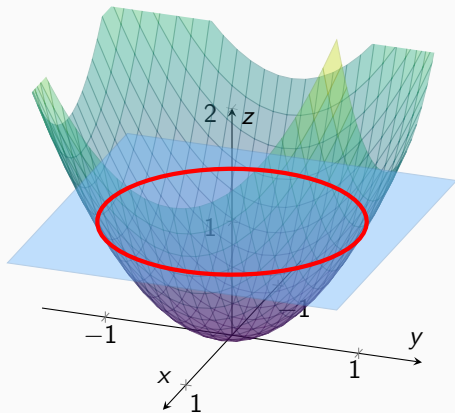
Tangent Space and Normal Space

Figure 1: Overlap between two equality constraints. The red line represents the overlapping region

Surface and curves (contd.)

The intersecting area (feasible space) in such equality constraints is a surface. In the figure we are dealing with constraints in 3D, so we get a 2D intersection region. Generally, we get a higher dimensional surface.

Definition (Curve)

A curve is a set of vectors $\boldsymbol{x}$ parametrized by some other variable t such that $t \in (a, b)$. It is usually represented by the set $\{\boldsymbol{x}(t)\}$. Curve is assumed to be differentiable $\forall t \in (a, b)$.

The constrained optimization problem has a solution if there exist an overlapping surface. If so, multiple curves can be possible on the surface.

Tangent space

On a particular curve $\mathbf{x}(t)$ such that for some $t^* \in (a, b) : \mathbf{x}^* = \mathbf{x}(t^*)$, $\dot{x}(t) = y$, we may define a $T(\mathbf{x}^*)$ which is called the tangent space.

Definition (Tangent space)

The Tangent space $T(\mathbf{x}^*) = \{\mathbf{y} \in \mathbb{R}^n | D\mathbf{h}(\mathbf{x}^*)\mathbf{y} = \mathbf{0}\}$.

- The above definition is equivalent to saying it is the set of tangents of all curves passing through $\mathbf{x}^*$. One may use this logic to state intuitively to say that if $T(\mathbf{x}^*) \neq \phi$ then there is at least one curve passing through $\mathbf{x}^*$.
- Tangent space is the null space of the constraint Jacobian. i.e.,

$$T(\mathbf{x}^*) = \mathcal{N}(D\mathbf{h}(\mathbf{x^*}))$$

- Geometrically, this tangent space is a hyperplane passing through origin and the vectors on this hyperplane are all orthogonal to all $\nabla h_i(\mathbf{x}^*) \forall i = 1, 2, \ldots, m$. If one transforms by adding $\mathbf{x^*} + \mathbf{y}$ where $pmby \in T(\mathbf{x}^*)$, the resulting set of such vectors $\mathbf{x^*} + \mathbf{y}$ is called the tangent plane.

Definition (Normal space)

The normal space is the set $N(x^*) = \{y \in \mathbb{R}^n | y = Dh(x^*)^T x, \forall x \in \mathbb{R}^m\}$

- $N(x^*) = \mathcal{R}(Dh(x^*)^T)$. i.e., the Normal space is the row space of the Jacobian.
- $N(x^*) = T(x^*)^\perp$. That is the Normal Space is orthogonal to tangent space.
- All constraint gradients $\nabla h_1(x), \nabla h_2(x), \ldots, \nabla h_m(x)$ are members of the Normal space.

FONC for equality constrained problems

Lagrange Multiplier theorem

Theorem (Lagrange multiplier theorem)

If x^* is a local maximizer and a regular point, then there exist a vector $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ such that,

$$\nabla f(x^*) - Dh(x^*)^T \boldsymbol{\lambda}^* = \mathbf{0}$$

What this theorem means is that if x^* happens to be a regular point and a local maximizer, then the objective function gradient at x^* can be expressed as a linear combination of the gradients of the constraints.

Geometrically,

- $\nabla f(x^*)$ lies in the space spanned by the gradients of the constraints $\nabla h_1(x^*), \nabla h_2(x^*), \dots, \nabla h_m(x^*)$. (In other words the Normal space)
- The gradient of the objective function will be orthogonal to the Tangent space at x^*. i.e., it will be perpendicular to the surface at x^*.

An example

Example

Maximize the area of a rectangular playground which can be constructed by a school. Safety regulations require that all playgrounds have a wall. The school has only a limited budget assigned for wall construction and cannot afford a wall of more than a certain length.

Sol. Let x_1 be the garden's length, x_2 its width. Let the fixed perimeter (length of wall) be $2P$. Also let $\boldsymbol{x}^* = [x_1^*, x_2^*]^T$ represent the optimal decision.

$$\max \ f(x_1, x_2) = x_1 x_2$$
$$\text{s.t.}, 2(x_1 + x_2) = 2P$$

Let's rewrite the constraint as $x_1 + x_2 - P = 0$. Since there is only one constraint, regaularity of $\boldsymbol{x}^*$ the Lagrange conditions are of the form, $\nabla f(\boldsymbol{x}^*) = \lambda \nabla h(\boldsymbol{x})$.

$$\nabla h(\boldsymbol{x}^*) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

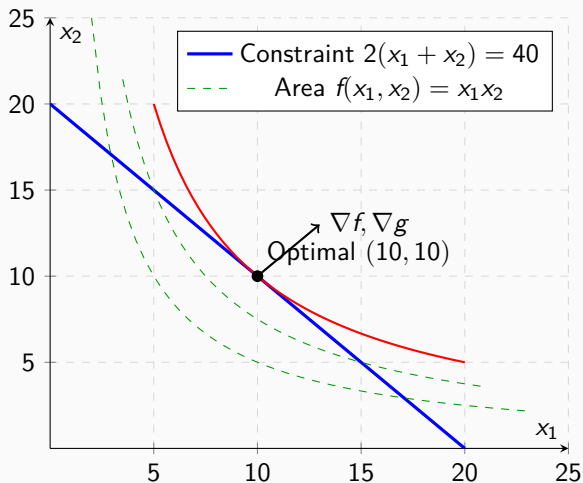$$\nabla h f(\boldsymbol{x}) = \begin{pmatrix} x_2^* \\ x_1^* \end{pmatrix}$$

Lagrange condition implies

$$\begin{pmatrix} x_2^* \\ x_1^* \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

i.e, $x_1^* = x_2^* = \frac{P}{2}$. Any potential maximizer must satisfy this requirement. Yet we cant be sure if it is in fact a maximizer.

Example (contd.)

Visualizing the Lagrange Condition

Proof of Lagrange multiplier theorem

TST, $\nabla f(\mathbf{x}^*) \in \mathcal{R}(D\mathbf{h}(\mathbf{x}^*)^T)$, i.e, the objective function gradient at $\mathbf{x}^*$ lies in the Normal space $N(\mathbf{x}^*)$.

- Consider any $\mathbf{y} \in T(\mathbf{x}^*)$. We know that there would now exist a curve $\mathbf{x}(t)$ such that $\mathbf{x}(t^*) = \mathbf{x}^*$ and $\dot{\mathbf{x}}(t^*) = \mathbf{y}$.
- If $\mathbf{x}^*$ is a maximizer, then unconstrained first order conditions with respect to t must hold at t^*.

$$\left(\frac{df(\mathbf{x}(t))}{dt} \right)_{t=t^*} = 0$$

$$\implies \nabla f(\mathbf{x}^*)^T \dot{\mathbf{x}}(t^*) = 0 \text{ By virtue of the chain rule}$$

$$\implies \nabla f(\mathbf{x}^*)^T \mathbf{y} = 0$$

That is a tangent vector at the maximizer is orthogonal to the gradient at the maximizer. Note that the above argument can be made for all $\mathbf{y} \in T(\mathbf{x}^*)$. i. e.,

$$\nabla f(\mathbf{x}^*)^T \mathbf{y} = 0 \qquad \forall \mathbf{y} \in T(\mathbf{x}^*)$$

$$\nabla f(\mathbf{x}^*) \in N(\mathbf{x}^*)$$