Non-Linear Programming

Constrained optimization: Equality and Inequality type constraints

Anand Jacob Abraham

Department of Industrial & Systems Engineering
IIT Kharagpur

Table of contents

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

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

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

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

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

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

Surface and curves

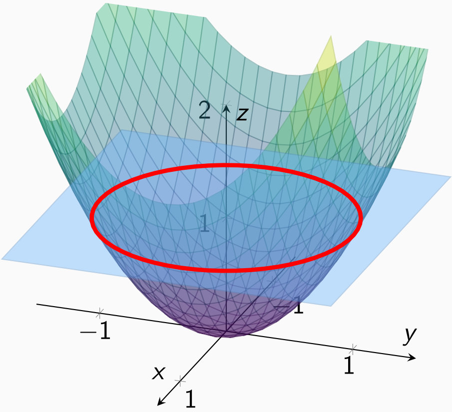


Figure 1: Overlap between two equality constraints. The red line represents the overlapping region

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

Definition (Tangent space)

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

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

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

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

Normal Space

Definition (Normal space)

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

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

FONC for equality constrained problems

Lagrange Multiplier theorem

Theorem (Lagrange multiplier theorem)

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

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

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

Geometrically,

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

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, regularity of $\boldsymbol{x}^*$ the Lagrange conditions are of the form, $\nabla f(\boldsymbol{x}^*) = \lambda \nabla h(\boldsymbol{x})$.

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

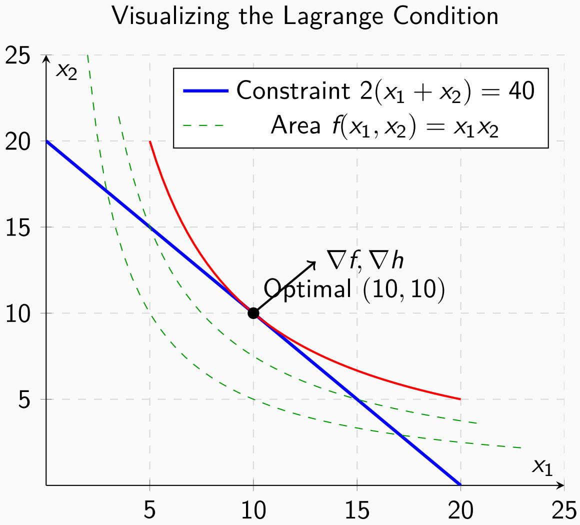
$$\nabla hf(\boldsymbol{x}) = \begin{pmatrix} x_2^* \\ x_1^* \end{pmatrix}$$

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.)



Proof of Lagrange multiplier theorem

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

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

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

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

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

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 $\boldsymbol{y} \in T(\boldsymbol{x}^*)$. i. e.,

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

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

Lagrange function

Definition (Lagrange function)

For an equality constrained maximization problem, the Lagrange function is given by $L(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) - \sum_{j=1}^{m} \lambda_j h_j(\boldsymbol{x})$.

The maximizers of L must satisfy the following requirement.

$$\left(\frac{\partial L}{\partial \lambda_j}\right)_{\boldsymbol{\lambda}=\boldsymbol{\lambda^*}, \boldsymbol{x}=\boldsymbol{x}^*} = 0 \qquad \forall j = 1, 2, \ldots, m \tag{1}$$

$$\left(\frac{\partial L}{\partial x_i}\right)_{\boldsymbol{\lambda}=\boldsymbol{\lambda^*}, \boldsymbol{x}=\boldsymbol{x}^*} = 0 \qquad \forall i = 1, 2, \ldots, n \tag{2}$$

- The FONC of unconstrained optimization when applied to the Lagrange function gives us the Lagrange conditions and the constraints respectively.
- $\therefore$The Lagrange function presents an unconstrained equivalent to our constrained optimization problem.

Economic interpretation of Lagrange multipliers

Consider the problem below.

$$\max \ f(\boldsymbol{x}) = x_1 + x_2$$
$$\text{s.t.} \ , x_1^2 + x_2^2 = 1$$

The solution to this problem is given below.

$$\nabla f\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) = \lambda^* \nabla h\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$$

$$\text{i.e.,} \ \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \lambda^* \begin{bmatrix} \sqrt{2} \\ \sqrt{2} \end{bmatrix}$$

Economic interpretation of Lagrange multipliers (contd.)

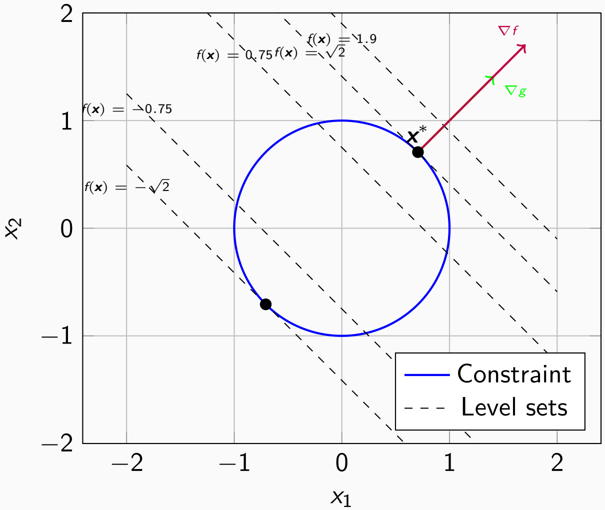


Figure 2: The constrained maximization of $x_1 + x_2$ subject to the constraint $g(\boldsymbol{x}) = x_1^2 + x_2^2 - 1 = 0$

Economic interpretation of Lagrange multipliers (contd.)

At $\boldsymbol{x^*} = [\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}]^T$, the value of the Lagrange multiplier $\lambda^* = \frac{1}{\sqrt{2}}$. Suppose, at this $\boldsymbol{x^*}$, we wish to find the change in objective function value when we make an incremental change δ to the constraint RHS. Assume $\delta > 0$ and $\delta \to 0$.

$$\frac{\partial f(\boldsymbol{x^*})}{\partial r} = \lambda^*.$$

Let the constraint be $x_1^2 + x_2^2 = r$. For $\delta > 0, \delta \to 0$, we get the following relationship. (For our problem $r = 1$.)

$$\frac{f(\boldsymbol{x^*})_{r+\delta} - f(\boldsymbol{x^*})_r}{\delta} = \lambda^*$$

Economic interpretation (contd.)

At $r = 1$, the expected change when we increase the RHS by δ unit is $\lambda^*\delta$. If the increase is by 1 unit, the expected change is $\lambda^* = \frac{1}{\sqrt{2}}$.

- This means λ^* may be interpreted as the maximum per unit price that the decision maker is willing to pay to increase the RHS r.
- This can be interpreted as her the decision maker's *Maximum Willingness to Pay (MWP)* or the *Shadow price* for the resource represented by the constraint.
- This information may aid the decision maker very much in a setting with multiple constraints.

Second Order Necessary Conditions (SONC)

Lemma (Second Order Necessary Condition)

If $\boldsymbol{x}^*$ is a regular point and a maximizer for $f(\boldsymbol{x})$ subject to constraints $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}$ such that $f, \boldsymbol{h} \in \mathcal{C}^2$, then there must exist $\boldsymbol{\lambda}^*$ which satisfies the following two conditions,

1. $\nabla f(\boldsymbol{x}^*) = D\boldsymbol{h}(\boldsymbol{x^*})^{\boldsymbol{T}}\boldsymbol{\lambda}^*$
2. $\boldsymbol{y}^T\boldsymbol{H_L}(\boldsymbol{x^*}, \boldsymbol{\lambda^*})\boldsymbol{y} \leq 0 \forall \boldsymbol{y} \in T(\boldsymbol{x}^*)$

- Here $\boldsymbol{H_L}(\boldsymbol{x^*}, \boldsymbol{\lambda^*})$ is the Hessian matrix for the Lagrange function (treating $\boldsymbol{x}$ as the decision variable) and evaluated at $\boldsymbol{x}^*, \boldsymbol{\lambda}^*$.
- You expect the Lagrange function to behave in a Negative semi-definite manner over the Tangent space.
- If any regular $\boldsymbol{x}^*$ violates the FONC or SONC, then it is definitely not a maximizer.

Second Order Sufficiency Conditions

Lemma (Second Order Sufficient Conditions)

For a maximization problem max $f(\boldsymbol{x})$ subject to $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}$ such that $f, \boldsymbol{h} \in \mathcal{C}^2$ and if there exist $\boldsymbol{x}^* \in \mathbb{R}^n$ and $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ which satisfies the following properties

1. $\nabla f(\boldsymbol{x}^*) = D\boldsymbol{h}(\boldsymbol{x^*})^{\boldsymbol{T}}\boldsymbol{\lambda^*}$
2. $\boldsymbol{y^T H_L(x^*, \lambda^*) y} < 0 \forall \boldsymbol{y} \in T(\boldsymbol{x}^*) : \boldsymbol{y} \neq \boldsymbol{0}$

then $\boldsymbol{x}^*$ is a strict local maximizer of f subject to $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}$

If a particular $\boldsymbol{x}^*$ satisfies FONC and SONC, but fails the SOSC, it may still be a maximizer.

Playground example revisited

Example

For the problem below, check if $\boldsymbol{x}^* = (\frac{P}{2}, \frac{P}{2})^T$ satisfies SONC and SOSC for local maxima.

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

Sol.

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = x_1 x_2 - \lambda(x_1 + x_2 - P)$$
$$\frac{\partial L}{\partial x_1} = x_2 - \lambda; \frac{\partial L}{\partial x_2} = x_1 - \lambda$$
$$\frac{\partial^2 L}{\partial x_1^2} = \frac{\partial^2 L}{\partial x_2^2} = 0$$
$$\frac{\partial^2 L}{\partial x_1 x_2} = 1$$
$$\boldsymbol{H}_L = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Playground example revisited (contd.)

$$
\begin{aligned}
T(\boldsymbol{x}^*) =& \{\boldsymbol{y} : [1\ 1][y_1\ \ y_2]^T = 0\} \\
=& \{\boldsymbol{y} : y_1 + y_2 = 0\}
\end{aligned}
$$

i.e., all tangents $\boldsymbol{y} = \begin{bmatrix} y_1 \\ -y_1 \end{bmatrix} : y_1 \in \mathbb{R}$.

$$
\begin{aligned}
\boldsymbol{y}^T \boldsymbol{H}_L \boldsymbol{y} =& [y_1\ \ -y_1] \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}] \begin{bmatrix} y_1 \\ -y_1 \end{bmatrix} \\
=& [y_1\ \ -y_1] \begin{bmatrix} -y_1 \\ y_1 \end{bmatrix} = -(y_1^2 + y_2^2) \leq 0 \forall \boldsymbol{y} \in T(\boldsymbol{x}^*)
\end{aligned}
$$

$\therefore$ SONC is satisfied. Also, $\forall \boldsymbol{y} \in T(\boldsymbol{x}^*) \setminus \{\boldsymbol{0}\}$, we have

$$
\boldsymbol{y}^T \boldsymbol{H}_L \boldsymbol{y} = -(y_1^2 + y_2^2) < 0 \quad \forall \boldsymbol{y} \in T(\boldsymbol{x}^*) \setminus \{\boldsymbol{0}\}
$$

$\therefore$ SOSC is satisfied at $\boldsymbol{x}^* = (\frac{P}{2}, \frac{P}{2})$. The point $\boldsymbol{x}^*$ is definitely a local maximizer.

Case study: Signal to Noise ratio

In communication applications, signal to noise ratio is a important parameter which needs to be optimized. Consider a radio device that receives signals from two towers located d distance apart. Currently tower 1 is transmitting the desired signal to the device and tower 2's signals act as interference. if the customer is located at a distance of l from tower 1, find the optimal value of l at which signal to noise ratio is maximized. Assume all towers are of height h. What is the expression for the optimal signal to noise ratio? Comment on what factors affect it?

Example

Examine if $\boldsymbol{x}^* = \mathbf{0}$ is a local maximizer to the following problem using the Lagrange consitions.

$$\max x_1^2$$

$$\text{Subject to,}$$

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

Yes. Change the objective function to $f(\boldsymbol{x}) = x_1$, explain what happens.

Optimization with inequality constraints

Problems with inequality constraints

Now the general structure of the problems that we deal with is the following.

$$\begin{aligned} \max\ & f(\boldsymbol{x}) \\ \text{s.t. } & \boldsymbol{h}(\boldsymbol{x}) = \mathbf{0} \\ & \boldsymbol{g}(\boldsymbol{x}) \leq \mathbf{0} \end{aligned}$$

where $\boldsymbol{h} : \mathbb{R}^n \mapsto \mathbb{R}^m$ and $\boldsymbol{g} : \mathbb{R}^n \mapsto \mathbb{R}^p$.

Definition (Binding constraint)

A constraint $g(\boldsymbol{x}) \leq 0$ is said to be binding or active at $\boldsymbol{x}_0$ if $g(\boldsymbol{x}_0) = 0$.

Let $B(\boldsymbol{x_0}) \equiv \{j : g_j(\boldsymbol{x}_0) = 0\}$.

Problems with inequality constraints (contd.)

Definition (Regular point)

A vector (point) $\boldsymbol{x}_0$ is called a regular point if $\nabla g_j(\boldsymbol{x}_{\mathbf{0}})$ are linearly independent $\forall j \in B(\boldsymbol{x}_0)$.

- So, the earlier definition of regularity is extended to encompass all binding constraints.
- All equality type constraints are by default binding.

KKT conditions

Theorem (KKT Conditions)

If $\boldsymbol{x^*}$ is a maximizer and a regular point, then $\exists \boldsymbol{\lambda^*} \in \mathbb{R}^m, \boldsymbol{\mu^*} \in \mathbb{R}^p$ is such that the following conditions hold.

1. $\nabla f(\boldsymbol{x^*}) = Dh(\boldsymbol{x^*})^T \boldsymbol{\lambda^*} + Dg(\boldsymbol{x}^*)^T \boldsymbol{\mu^*}$.
2. $\mu_j^* g_j(\boldsymbol{x^*}) = 0 \forall \quad j \in \{1, 2, \ldots, p\}$
3. $\boldsymbol{\mu^*} \geq \mathbf{0}$.

In optimization literature $\boldsymbol{\mu}$ is also called the *KKT multiplier vector*.

Example

Example (Exercise)

Show that if $\boldsymbol{x}^*$ is a maximizer and $\exists \boldsymbol{\lambda^*} \in \mathbb{R}^m$ which satisfies $\nabla f(\boldsymbol{x}^*) = Dh(\boldsymbol{x}^*)^T \boldsymbol{\lambda^*}$, then $\boldsymbol{x}^* \in \mathbb{R}^n, \boldsymbol{\lambda^*} \in \mathbb{R}^n$ must satisfy the FONC for the following unconstrained optimization problem below.

$$\max_{\boldsymbol{x},\boldsymbol{\lambda}} L(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) - \boldsymbol{\lambda}^T \boldsymbol{g}(\boldsymbol{x})$$

Lagrangian for inequality constraints

Similar to the equality constrained setting we can define the Lagrangian as follows.

$$L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\boldsymbol{x}) - \sum_{k=1}^{m} \lambda_k h_k(\boldsymbol{x}) - \sum_{j=1}^{p} \mu_j g_j(\boldsymbol{x})$$

If there is some $\boldsymbol{x}^*$ which is a local maximizer and some $\boldsymbol{\lambda^*} \in \mathbb{R}^m$ and $\boldsymbol{\mu^*} \in \mathbb{R}^p$ which satisfy the first KKT condition, then $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ should be unconstrained maximizers of $L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$.

Second Order Necessary Conditions

Lemma (SONC)

If $\boldsymbol{x}^*$ is a regular point and a maximizer, it must satisfy $\boldsymbol{y}^T H_L(\boldsymbol{x}^*, \boldsymbol{\lambda^*}, \boldsymbol{\mu^*})\boldsymbol{y} \leq 0 \forall \boldsymbol{y} \in T(\boldsymbol{x}^*)$. Here $T(\boldsymbol{x}^*)$ is the tangent space at $\boldsymbol{x}^*$.

Proof.

$T(\boldsymbol{x}^*)$ is the tangent space at $\boldsymbol{x}^*$ on the surface formed by active constraints. So all active constraints are equalities and therefore the SONC for local maximizers of equality constraints should hold true. □

Second Order Sufficient Conditions

Definition (Relaxed tangent space)

The relaxed tangent space $T'(\boldsymbol{x}^*)$ at the stationary point is defined in the following manner.

$$T'(\boldsymbol{x}^*) \equiv \{\boldsymbol{y} : D\boldsymbol{h}(\boldsymbol{x})\boldsymbol{y} = \boldsymbol{0}, \boldsymbol{y}^T \nabla g_j(\boldsymbol{x}^*) = 0 \forall j \in B(\boldsymbol{x}^*) : \mu_j^* > 0\}$$

Lemma

[SOSC] if $\boldsymbol{x}^*$ is a feasible point which satisfies the following conditions

1. *KKT conditions for maximizers.*
2. $\boldsymbol{y}^T H_L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)\boldsymbol{y} < 0 \forall \boldsymbol{y} \in T'(\boldsymbol{x}^*) \setminus \boldsymbol{0}$

then it is a local maximizer.

- The tangent space is a subset of the relaxed Tangent Space. i.e., $T(\boldsymbol{x}) \subseteq T'(\boldsymbol{x})$.
- SOSC requires the matrix $\boldsymbol{H_L}$ to behave negative definitely over a larger space as compared to before. This is a much stronger condition than what was required for SOSC of equality constraints.