MATH 4211/6211 – Optimization

Constrained Optimization

Xiaojing Ye
Department of Mathematics & Statistics
Georgia State University

Constrained optimization problems are formulated as

$$\begin{aligned} \text{minimize} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & g_j(\boldsymbol{x}) \le 0, \quad j = 1, \dots, p, \\ & h_i(\boldsymbol{x}) = 0, \quad i = 1, \dots, m, \end{aligned}$$

where $g_j, h_i : \mathbb{R}^n \to \mathbb{R}$ are **inequality** and **equality** constraint functions, respectively.

We can summarize them into vector-valued functions $\boldsymbol{g}$ and $\boldsymbol{h}$:

$$\boldsymbol{g}(\boldsymbol{x}) = \begin{bmatrix} g_1(\boldsymbol{x}) \\ \vdots \\ g_p(\boldsymbol{x}) \end{bmatrix} \quad \text{and} \quad \boldsymbol{h}(\boldsymbol{x}) = \begin{bmatrix} h_1(\boldsymbol{x}) \\ \vdots \\ h_m(\boldsymbol{x}) \end{bmatrix}$$

so the constraints can be written as $\boldsymbol{g}(\boldsymbol{x}) \le \boldsymbol{0}$ and $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}$, respectively.

Note that $\boldsymbol{g} : \mathbb{R}^n \to \mathbb{R}^p$ and $\boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^m$.

We can write the constrained optimization concisely as

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

The feasible set is $\Omega := \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{g}(\boldsymbol{x}) \le \boldsymbol{0}, \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}\}$.

Exmaple. LP (standard form) is a constrained optimization with $f(\boldsymbol{x}) = \boldsymbol{c}^\top \boldsymbol{x}$, $\boldsymbol{g}(\boldsymbol{x}) = -\boldsymbol{x}$ and $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}$.

We now focus on constrained optimization problems with equality constraints only, i.e.,

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

and the feasible set is $\Omega = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}\}$.

Some equality-constrained optimization problem can be converted into unconstrained ones.

Example.

- Consider the constrained optimization problem

$$\begin{aligned} \text{minimize} \quad & x_1^2 + 2x_1x_2 + 3x_2^2 + 4x_1 + 5x_2 + 6x_3 \\ \text{subject to} \quad & x_1 + 2x_2 = 3 \\ & 4x_1 + 5x_3 = 6 \end{aligned}$$

  The constraints imply that $x_2 = \frac{1}{2}(3 - x_1)$ and $x_3 = \frac{1}{5}(6 - x_1)$. Substitute x_2 and x_3 in the objective function to get an unconstrained minimization of x_1 only.

Example.

- Consider the constrained optimization problem

$$\begin{aligned} \text{maximize} \quad & x_1 x_2 \\ \text{subject to} \quad & x_1^2 + 4x_2^2 = 1 \end{aligned}$$

  It is equivalent to maximizing $x_1^2 x_2^2$ then substitute x_1^2 by $1 - 4x_2^2$ to get an unconstrained problem of x_2.

  Another way to solving this is using $1 = x_1^2 + (2x_2)^2 \geq 4x_1x_2$ where the equality holds when $x_1 = 2x_2$.So $x_1 = \sqrt{2}/2$ and $x_2 = \sqrt{2}/4$.

However, not all equality-constrained problems can be easily converted into unconstrained ones.

We need general theory to solve constrained optimization problems with equality constraints:

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

and the feasible set is $\Omega = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}\}$.

Recall that $\boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^m$ ($m \le n$) has Jacobian matrix

$$D\boldsymbol{h}(\boldsymbol{x}) = \begin{bmatrix} \nabla h_1(\boldsymbol{x})^\top \\ \vdots \\ \nabla h_m(\boldsymbol{x})^\top \end{bmatrix} \in \mathbb{R}^{m \times n}$$

Definition. We say a point $\boldsymbol{x} \in \Omega$ is a **regular point** if $\text{rank}(D\boldsymbol{h}(\boldsymbol{x})) = m$, i.e., the Jacobian matrix has full row rank.

Example. Let $n = 3$ and $m = 1$. Define $h_1(\boldsymbol{x}) = x_2 - x_3^2$ be the only constraint. Then the Jacobian matrix is

$$D\boldsymbol{h}(\boldsymbol{x}) = [\nabla h_1(\boldsymbol{x})^\top] = [0, 1, -2x_3]$$

Note that $D\boldsymbol{h}(\boldsymbol{x}) \neq \mathbf{0}$ and hence rank$(D\boldsymbol{h}(\boldsymbol{x})) = 1$ everywhere.

The feasible set Ω is a “surface” in $\mathbb{R}^3$ with dimension $n - m = 3 - 1 = 2$.

Example. Let $n = 3$ and $m = 2$. Define $h_1(\boldsymbol{x}) = x_1$ and $h_2(\boldsymbol{x}) = x_2 - x_3^2$. The Jacobian is

$$D\boldsymbol{h}(\boldsymbol{x}) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -2x_3 \end{bmatrix}$$

with $\text{rank}(D\boldsymbol{h}(\boldsymbol{x})) = 2$ everywhere, and the feasible set Ω is a line in $\mathbb{R}^3$ with dimension $n - m = 3 - 2 = 1$.

Tangent space and normal space

Defintion. We say $\boldsymbol{x} : (a, b) \to \mathbb{R}^n$, a curve in $\mathbb{R}^n$, is **differentiable** if $x_i'(t)$ exists for all $t \in (a, b)$. The derivative is defined by

$$\boldsymbol{x}'(t) = \begin{bmatrix} x_1'(t) \\ \vdots \\ x_n'(t) \end{bmatrix}$$

We say $\boldsymbol{x}$ is **twice differentiable** if $x_i''(t)$ exists for all $t \in (a, b)$, and

$$\boldsymbol{x}''(t) = \begin{bmatrix} x_1''(t) \\ \vdots \\ x_n''(t) \end{bmatrix}$$

Defintion. The **tangent space** of $\Omega = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}\}$ at $\boldsymbol{x}^*$ is the set

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

In other words, $T(\boldsymbol{x}^*) = \mathcal{N}(D\boldsymbol{h}(\boldsymbol{x}^*))$.

Remark. If $\boldsymbol{x}^*$ is regular, then rank$(D\boldsymbol{h}(\boldsymbol{x}^*)) = m$, and hence $\dim(T(\boldsymbol{x}^*)) = \dim(\mathcal{N}(D\boldsymbol{h}(\boldsymbol{x}^*))) = n - m$.

Remark. We sometimes draw the tangent space as a plane tangent to Ω at $\boldsymbol{x}^*$, that tangent plane is

$$TP(\boldsymbol{x}^*) := \boldsymbol{x}^* + T(\boldsymbol{x}^*) = \{\boldsymbol{x}^* + \boldsymbol{y} : \boldsymbol{y} \in T(\boldsymbol{x}^*)\}$$

Example. Let

$$\Omega = \{\boldsymbol{x} \in \mathbb{R}^3 : h_1(\boldsymbol{x}) = x_1 = 0,\ h_2(\boldsymbol{x}) = x_1 - x_2 = 0\}$$

Then we have

$$D\boldsymbol{h}(\boldsymbol{x}) = \begin{bmatrix} \nabla h_1(\boldsymbol{x})^\top \\ \nabla h_2(\boldsymbol{x})^\top \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & -1 & 0 \end{bmatrix}$$

at any $\boldsymbol{x} \in \Omega$, and the tangent space at any point $\boldsymbol{x}$ is

$$\begin{aligned} T(\boldsymbol{x}) = \mathcal{N}(D(\boldsymbol{h}(\boldsymbol{x}))) &= \left\{ \boldsymbol{y} \in \mathbb{R}^3 : \begin{bmatrix} 1 & 0 & 0 \\ 1 & -1 & 0 \end{bmatrix} \boldsymbol{y} = \boldsymbol{0} \right\} \\ &= \{[0; 0; \alpha] \in \mathbb{R}^3 : \alpha \in \mathbb{R}\} \end{aligned}$$

Theorem. Suppose $\boldsymbol{x}^*$ is regular. Then $\boldsymbol{y} \in T(\boldsymbol{x}^*)$ iff there exists curve $\boldsymbol{x} : (-\delta, \delta) \to \Omega$ such that $\boldsymbol{x}(0) = \boldsymbol{x}^*$ and $\boldsymbol{x}'(0) = \boldsymbol{y}$.

Proof. ($\Leftarrow$) Let $\boldsymbol{x}(t)$ be such a curve, then $\boldsymbol{h}(\boldsymbol{x}(t)) = \mathbf{0}$ for $t \in (-\delta, \delta)$ and

$$D\boldsymbol{h}(\boldsymbol{x}(0))\boldsymbol{x}'(0) = D\boldsymbol{h}(\boldsymbol{x}^*)\boldsymbol{y} = 0$$

which implies $\boldsymbol{y} \in T(\boldsymbol{x}^*)$.

Proof (cont.)

($\Rightarrow$) For any t, let $\boldsymbol{u} = \boldsymbol{u}(t) \in \mathbb{R}^n$ be determined by t s.t. it solves

$$\bar{\boldsymbol{h}}(t, \boldsymbol{u}) := \boldsymbol{h}(\boldsymbol{x}^* + t\boldsymbol{y} + D\boldsymbol{h}(\boldsymbol{x}^*)^\top \boldsymbol{u}) = \mathbf{0}$$

We know $\boldsymbol{u}(0) = \mathbf{0}$ is a solution at $t = 0$. Moreover,

$$D_{\boldsymbol{u}}\bar{\boldsymbol{h}}(0, \boldsymbol{u}) = D\boldsymbol{h}(\boldsymbol{x}^*)D\boldsymbol{h}(\boldsymbol{x}^*)^\top \succ \mathbf{0}$$

as $D\boldsymbol{h}(\boldsymbol{x}^*)$ has full row rank. Hence by Implicit Function Theorem, there is $\delta > 0$ s.t. a unique solution $\boldsymbol{u}(t)$ to $\bar{\boldsymbol{h}}(t, \boldsymbol{u}) = \mathbf{0}$ exists for $t \in (-\delta, \delta)$. Then

$$\boldsymbol{x}(t) = \boldsymbol{x}^* + t\boldsymbol{y} + D\boldsymbol{h}(\boldsymbol{x}^*)^\top \boldsymbol{u}(t)$$

is the desired curve.

Defintion. The normal space of Ω at $\boldsymbol{x}^*$ is defined by

$$N(\boldsymbol{x}^*) = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{x} = D\boldsymbol{h}(\boldsymbol{x}^*)^\top \boldsymbol{z} \text{ for some } \boldsymbol{z} \in \mathbb{R}^m\}$$

In other words,

$$\begin{aligned} N(\boldsymbol{x}^*) &= \mathcal{C}(D\boldsymbol{h}(\boldsymbol{x}^*)^\top) \\ &= \mathcal{R}(D\boldsymbol{h}(\boldsymbol{x}^*)) \\ &= \text{span}\{\nabla h_1(\boldsymbol{x}^*), \dots, \nabla h_m(\boldsymbol{x}^*)\} \end{aligned}$$

Note that $\text{dim}(N(\boldsymbol{x}^*)) = \text{dim}(\mathcal{R}(D\boldsymbol{h}(\boldsymbol{x}^*))) = m$.

Remark. The tangent space $T(\boldsymbol{x}^*)$ and the normal space $N(\boldsymbol{x}^*)$ form an orthogonal decomposition of $\mathbb{R}^n$:

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

where $T(\boldsymbol{x}^*) \perp N(\boldsymbol{x}^*)$.

We can also write this as $T(\boldsymbol{x}^*)^\perp = N(\boldsymbol{x}^*)$ or $N(\boldsymbol{x}^*)^\perp = T(\boldsymbol{x}^*)$.

Hence, for any $\boldsymbol{v} \in \mathbb{R}^n$, there exist a unique pair $\boldsymbol{y} \in T(\boldsymbol{x}^*)$ and $\boldsymbol{w} \in N(\boldsymbol{x}^*)$, such that

$$\boldsymbol{v} = \boldsymbol{y} + \boldsymbol{w}$$

Now let us see the first-order necessary conditions (FONC) for equality-constrained minimization.

Suppose $\boldsymbol{x}^*$ is a local minimizer of $f(\boldsymbol{x})$ over $\Omega = \{\boldsymbol{x} : \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}\}$, where $f, \boldsymbol{h} \in C^1$.

Then *for any* $\boldsymbol{y} \in T(\boldsymbol{x}^*)$, there exists curve $\boldsymbol{x} : (a, b) \to \Omega$ such that $\boldsymbol{x}(t) = \boldsymbol{x}^*$ and $\boldsymbol{x}'(t) = \boldsymbol{y}$ for some $t \in (a, b)$.

Define $\phi(s) = f(\boldsymbol{x}(s))$ (note that $\phi : (a, b) \to \mathbb{R}$), then

$$\phi'(s) = \nabla f(\boldsymbol{x}(s))^\top \boldsymbol{x}'(s)$$

In particular, due to the standard FONC, we have

$$\phi'(t) = \nabla f(\boldsymbol{x}(t))^\top \boldsymbol{x}'(t) = \nabla f(\boldsymbol{x}^*)^\top \boldsymbol{y} = 0$$

Since $\boldsymbol{y} \in T(\boldsymbol{x}^*)$ is arbitrary, we know $\nabla f(\boldsymbol{x}^*) \perp T(\boldsymbol{x}^*)$, i.e.,

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

This means that $\exists\, \boldsymbol{\lambda}^* \in \mathbb{R}^m$, such that

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

This result is summarized below:

Theorem [Lagrange's Theorem]. If x^* is a local minimizer (or maximizer) of $f : \mathbb{R}^n \to \mathbb{R}$ subject to $\boldsymbol{h}(\boldsymbol{x}) = \mathbf{0} \in \mathbb{R}^m$ where $m \le n$, and $\boldsymbol{x}^*$ is a regular point ($D\boldsymbol{h}(\boldsymbol{x}^*)$ has full row rank), then there exists $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ s.t.

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

Now we know if $\boldsymbol{x}^*$ is a local minimizer of

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

then $\boldsymbol{x}^*$ must satisfy

$$\begin{aligned} \nabla f(\boldsymbol{x}^*) + D\boldsymbol{h}(\boldsymbol{x}^*)^\top \boldsymbol{\lambda}^* = \boldsymbol{0} \\ \boldsymbol{h}(\boldsymbol{x}^*) = \boldsymbol{0} \end{aligned}$$

There are called the first-order necessary conditions (FNOC), or the **Lagrange condition**, of the equality-constrained minimization problem. $\boldsymbol{\lambda}^*$ is called the **Lagrange multiplier**.

Remark. The conditions above are necessary but not sufficient to determine $\boldsymbol{x}^*$ to be a local minimizer—a point satisfying these conditions could be a local maximizer or neither.

Example. Consider the problem

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

where $f(x) = x$ and

$$h(x) = \begin{cases} x^2 & \text{if } x < 0 \\ 0 & \text{if } 0 \le x \le 1 \\ (x-1)^2 & \text{if } x > 1 \end{cases}$$

We can see that $\Omega = [0, 1]$ and $x^* = 0$ is the only local minimizer. However $f'(x^*) = 1$ and $h'(x^*) = 0$. The Lagrange condition fails to hold because x^* is not a regular point.

We introduce the **Lagrange function**

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

Then the Lagrange condition becomes

$$\begin{aligned} \nabla_{\boldsymbol{x}} l(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) &= \nabla f(\boldsymbol{x}^*) + D\boldsymbol{h}(\boldsymbol{x}^*)^\top \boldsymbol{\lambda}^* = \mathbf{0} \\ \nabla_{\boldsymbol{\lambda}} l(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) &= \boldsymbol{h}(\boldsymbol{x}^*) = \mathbf{0} \end{aligned}$$

Note that this is a system of $n + m$ equations for $[\boldsymbol{x}; \boldsymbol{\lambda}] \in \mathbb{R}^{n+m}$ (which as $n + m$ unknowns).

Example. Given a fixed area A of cardboard, we wish to construct a closed cardboard box with maximum volume. Let the dimension of the box be $\boldsymbol{x} = [x_1; x_2; x_3]$, then the problem can be formulated as

$$\begin{aligned} \text{maximize} \quad & x_1 x_2 x_3 \\ \text{subject to} \quad & 2(x_1 x_2 + x_2 x_3 + x_3 x_1) = A \end{aligned}$$

Hence we can set

$$\begin{aligned} f(\boldsymbol{x}) &= -x_1 x_2 x_3 \\ h(\boldsymbol{x}) &= x_1 x_2 + x_2 x_3 + x_3 x_1 - \frac{A}{2} \end{aligned}$$

Then the Lagrange function is

$$l(\boldsymbol{x}, \lambda) = f(\boldsymbol{x}) + h(\boldsymbol{x})\lambda = -x_1 x_2 x_3 + \left(x_1 x_2 + x_2 x_3 + x_3 x_1 - \frac{A}{2} \right) \lambda$$

So the Lagrange condition is

$$\nabla_x l(\boldsymbol{x}, \lambda) = 0$$
$$\nabla_\lambda l(\boldsymbol{x}, \lambda) = 0$$

which is

$$x_2 x_3 - (x_2 + x_3)\lambda = 0$$
$$x_1 x_3 - (x_1 + x_3)\lambda = 0$$
$$x_1 x_2 - (x_1 + x_2)\lambda = 0$$
$$x_1 x_2 + x_2 x_3 + x_3 x_1 - \frac{A}{2} = 0$$

Then solving this system yields

$$x_1 = x_2 = x_3 = \sqrt{\frac{A}{6}}, \quad \lambda = \frac{1}{2}\sqrt{\frac{A}{6}}.$$

Example. Consider an equality-constrained optimization problem

$$\begin{aligned} \text{minimize} \quad & x_1^2 + x_2^2 \\ \text{subject to} \quad & x_1^2 + 2x_2^2 = 1 \end{aligned}$$

Solution. Here $f(\boldsymbol{x}) = x_1^2 + x_2^2$ and $\boldsymbol{h}(\boldsymbol{x}) = x_1^2 + 2x_2^2 - 1$.

The Lagrange function is

$$l(\boldsymbol{x}, \boldsymbol{\lambda}) = (x_1^2 + x_2^2) + \lambda(x_1^2 + 2x_2^2 - 1)$$

Then we obtain

$$\begin{aligned} \partial_{x_1} l(\boldsymbol{x}, \lambda) &= 2x_1 + 2\lambda x_1 = 0 \\ \partial_{x_2} l(\boldsymbol{x}, \lambda) &= 2x_2 + 4\lambda x_2 = 0 \\ \partial_{\lambda} l(\boldsymbol{x}, \lambda) &= x_1^2 + 2x_2^2 - 1 = 0 \end{aligned}$$

Solution (cont). Solving this system yields

$$\begin{bmatrix} x_1 \\ x_2 \\ \lambda \end{bmatrix} = \begin{bmatrix} 0 \\ 1/\sqrt{2} \\ -1/2 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ -1/\sqrt{2} \\ -1/2 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \quad \text{and} \begin{bmatrix} -1 \\ 0 \\ -1 \end{bmatrix}$$

It is easy to check that $\boldsymbol{x} = [0; \pm 1/\sqrt{2}]$ are local minimizers, and $\boldsymbol{x} = [0; \pm 1]$ are local maximizers.

Now we consider second-order conditions. We assume $f, \boldsymbol{h} \in C^2$.

Following the same steps as in FONC, suppose $\boldsymbol{x}^*$ is a local minimizer, then for any $\boldsymbol{y} \in T(\boldsymbol{x}^*)$, there exists a curve $\boldsymbol{x} : (a, b) \to \Omega$ such that $\boldsymbol{x}(t) = \boldsymbol{x}^*$ and $\boldsymbol{x}'(t) = \boldsymbol{y}$ for some $t \in (a, b)$.

Again define $\phi(s) = f(\boldsymbol{x}(s))$, and hence $\phi'(s) = \nabla f(\boldsymbol{x}(s))^\top \boldsymbol{x}'(s)$. Then the standard second-order necessary condition (SONC) implies that at a local minimizer there are

$$\phi'(t) = \nabla f(\boldsymbol{x}(t))^\top \boldsymbol{x}'(t) = \nabla f(\boldsymbol{x}^*)^\top \boldsymbol{y} = \mathbf{0}$$

and

$$\phi''(t) = \boldsymbol{y}^\top \nabla^2 f(\boldsymbol{x}^*) \boldsymbol{y} + \nabla f(\boldsymbol{x}^*)^\top \boldsymbol{x}''(t) \geq 0$$

In addition, since $\psi_i(s) := h_i(\boldsymbol{x}(s)) = 0$ for all $s \in (a, b)$, we have $\psi_i''(t) = 0$ which yields

$$\boldsymbol{y}^\top \nabla^2 h_i(\boldsymbol{x}^*)\boldsymbol{y} + \nabla h_i(\boldsymbol{x}^*)^\top \boldsymbol{x}''(t) = 0$$

for all $i = 1, \dots, m$.

According to the Lagrange condition, we know $\exists\, \boldsymbol{\lambda}^* \in \mathbb{R}^m$ such that

$$\nabla f(\boldsymbol{x}^*) + D\boldsymbol{h}(\boldsymbol{x}^*)^\top \boldsymbol{\lambda}^* = \nabla f(\boldsymbol{x}^*) + \sum_{i=1}^m \lambda_i^* \nabla h_i(\boldsymbol{x}^*) = \boldsymbol{0}$$

Using the results above, we can cancel the term with $\boldsymbol{x}''(t)$ and obtain

$$\boldsymbol{y}^\top \left[\nabla^2 f(\boldsymbol{x}^*) + \sum_{i=1}^m \lambda_i^* \nabla^2 h_i(\boldsymbol{x}^*)\right] \boldsymbol{y} \geq 0$$

for all $\boldsymbol{y} \in T(\boldsymbol{x}^*)$.

We summarize the second-order necessary condition (SONC):

Theorem (SONC). Let $\boldsymbol{x}^*$ be a local minimizer of $f : \mathbb{R}^n \to \mathbb{R}$ over $\Omega = \{\boldsymbol{x} : \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0} \in \mathbb{R}^m\}$ with $m \le n$, where $f, \boldsymbol{h} \in C^2$. Suppose $\boldsymbol{x}^*$ is regular, then $\exists\, \boldsymbol{\lambda}^* = [\lambda_1^*; \dots; \lambda_m^*] \in \mathbb{R}^m$ such that

1. $\nabla f(\boldsymbol{x}^*) + D\boldsymbol{h}(\boldsymbol{x}^*)^\top \boldsymbol{\lambda}^* = \boldsymbol{0}$;
2. For every $\boldsymbol{y} \in T(\boldsymbol{x}^*)$, there is

$$\boldsymbol{y}^\top \left[\nabla^2 f(\boldsymbol{x}^*) + \sum_{i=1}^m \lambda_i^* \nabla^2 h_i(\boldsymbol{x}^*)\right] \boldsymbol{y} \ge 0.$$

So $\nabla^2 f(\boldsymbol{x}^*) + \sum_{i=1}^m \lambda_i^* \nabla^2 h_i(\boldsymbol{x}^*)$ is playing the role of “Hessian”.

We also have the following second-order sufficient condition (SOSC):

Theorem (SOSC). Suppose $\boldsymbol{x}^* \in \Omega = \{\boldsymbol{x} : \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}\}$ is regular. If $\exists\, \boldsymbol{\lambda}^* = [\lambda_1^*; \dots; \lambda_m^*] \in \mathbb{R}^m$ such that

1. $\nabla f(\boldsymbol{x}^*) + D\boldsymbol{h}(\boldsymbol{x}^*)^\top \boldsymbol{\lambda}^* = \boldsymbol{0}$;
2. for every nonzero $\boldsymbol{y} \in T(\boldsymbol{x}^*)$, there is

$$\boldsymbol{y}^\top \left[\nabla^2 f(\boldsymbol{x}^*) + \sum_{i=1}^m \lambda_i^* \nabla^2 h_i(\boldsymbol{x}^*) \right] \boldsymbol{y} > 0.$$

Then $\boldsymbol{x}^*$ is a strict local minimizer of f over Ω.

Example. Solve the following problem

$$\text{maximize} \quad \frac{\boldsymbol{x}^\top \boldsymbol{Q} \boldsymbol{x}}{\boldsymbol{x}^\top \boldsymbol{P} \boldsymbol{x}}$$

where $\boldsymbol{Q} = \text{diag}([4, 1])$ and $\boldsymbol{P} = \text{diag}([2, 1])$.

Solution. Note the objective function is scale-invariant (replacing $\boldsymbol{x}$ by $t\boldsymbol{x}$ for any $t \neq 0$ yields the same value). This can be converted into the constrained minimization problem

$$\begin{aligned} \text{minimize} \quad & -\boldsymbol{x}^\top \boldsymbol{Q} \boldsymbol{x} \\ \text{subject to} \quad & \boldsymbol{x}^\top \boldsymbol{P} \boldsymbol{x} - 1 = 0 \end{aligned}$$

and $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{P} \boldsymbol{x} - 1 \in \mathbb{R}$ is the constraint.

Note that $D\boldsymbol{h}(\boldsymbol{x}) = 2\boldsymbol{P}\boldsymbol{x} = [4x_1; 2x_2]$.

Solution (cont). We first write the Lagrange function

$$l(\boldsymbol{x}, \lambda) = -\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} + \lambda(\boldsymbol{x}^\top \boldsymbol{P}\boldsymbol{x} - 1)$$

Then the Lagrange condition becomes

$$\nabla_{\boldsymbol{x}} l(\boldsymbol{x}^*, \lambda^*) = -2(\boldsymbol{Q} - \lambda^* \boldsymbol{P})\boldsymbol{x}^* = \boldsymbol{0}$$
$$\nabla_{\boldsymbol{\lambda}} l(\boldsymbol{x}^*, \lambda^*) = \boldsymbol{x}^{*\top} \boldsymbol{P}\boldsymbol{x}^* - 1 = \boldsymbol{0}$$

The first equation implies $\boldsymbol{P}^{-1}\boldsymbol{Q}\boldsymbol{x}^* = \lambda^* \boldsymbol{x}^*$, and hence λ^* is an eigenvalue of $\boldsymbol{P}^{-1}\boldsymbol{Q} = \text{diag}([2, 1])$. Hence $\lambda^* = 2$ or $\lambda^* = 1$.

For $\lambda^* = 2$, we know $\boldsymbol{x}^*$ is the corresponding eigenvector of $\boldsymbol{P}^{-1}\boldsymbol{Q}$ and satisfies ${\boldsymbol{x}^*}^\top \boldsymbol{P}\boldsymbol{x}^* = 1$. Hence $\boldsymbol{x}^* = [\pm 1/\sqrt{2}; 0]$. The tangent space is $T(\boldsymbol{x}^*) = \mathcal{N}(D\boldsymbol{h}(\boldsymbol{x}^*)) = \mathcal{N}([\pm\sqrt{2}; 0]) = \{[0; a] : a \in \mathbb{R}\}$.

We also have

$$\nabla^2 f(\boldsymbol{x}^*) + \lambda^* \nabla^2 \boldsymbol{h}(\boldsymbol{x}^*) = -2\boldsymbol{Q} + 2\lambda^* \boldsymbol{P} = \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix}$$

Therefore $\boldsymbol{y}^\top [\nabla^2 f(\boldsymbol{x}^*) + \lambda^* \nabla^2 \boldsymbol{h}(\boldsymbol{x}^*)]\boldsymbol{y} = 2a^2 > 0$ for all $\boldsymbol{y} = [0; a] \in T(\boldsymbol{x}^*)$ with $a \neq 0$.

Therefore $\boldsymbol{x}^* = [\pm 1/\sqrt{2}; 0]$ are both strict local minimizers of the constrained optimization problem.

Going back to the original problem, any $\boldsymbol{x}^* = [t; 0]$ with $t \neq 0$ is a strict local maximizer of $\frac{\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x}}{\boldsymbol{x}^\top \boldsymbol{P}\boldsymbol{x}}$.

For $\lambda^* = 1$, we know $\boldsymbol{x}^*$ is the corresponding eigenvector of $\boldsymbol{P}^{-1}\boldsymbol{Q}$ and satisfies ${\boldsymbol{x}^*}^\top \boldsymbol{P}\boldsymbol{x}^* = 1$. Hence $\boldsymbol{x}^* = [0; \pm 1]$. The tangent space is $T(\boldsymbol{x}^*) = \mathcal{N}(D\boldsymbol{h}(\boldsymbol{x}^*)) = \mathcal{N}([0; \pm 1]) = \{[a; 0] : a \in \mathbb{R}\}$.

We also have

$$\nabla^2 f(\boldsymbol{x}^*) + \lambda^* \nabla^2 \boldsymbol{h}(\boldsymbol{x}^*) = -2\boldsymbol{Q} + 2\lambda^* \boldsymbol{P} = \begin{bmatrix} -4 & 0 \\ 0 & 0 \end{bmatrix}$$

Therefore $\boldsymbol{y}^\top[\nabla^2 f(\boldsymbol{x}^*) + \lambda^* \nabla^2 \boldsymbol{h}(\boldsymbol{x}^*)]\boldsymbol{y} = -4a^2 < 0$ for all $\boldsymbol{y} = [a; 0] \in T(\boldsymbol{x}^*)$ with $a \neq 0$.

Therefore $\boldsymbol{x}^* = [0; \pm 1]$ are both strict local maximizers of the constrained optimization problem.

Going back to the original problem, any $\boldsymbol{x}^* = [0; t]$ with $t \neq 0$ is strict local minimizer of $\frac{\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x}}{\boldsymbol{x}^\top \boldsymbol{P}\boldsymbol{x}}$.

Now we consider a special type of constrained minimization problem with linear equality constraints (again $\boldsymbol{Q} \succ \mathbf{0}$ and $\boldsymbol{A}$ has full row rank):

$$\begin{array}{rl} \text{minimize} & \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} \\ \text{subject to} & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{array}$$

We have $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x}$ and $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}$.

The Lagrange function is

$$l(\boldsymbol{x}, \boldsymbol{\lambda}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{\lambda}^\top (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}).$$

Hence the Lagrange condition is

$$\begin{aligned} \nabla_{\boldsymbol{x}} l(\boldsymbol{x}, \boldsymbol{\lambda}) &= \boldsymbol{Q}\boldsymbol{x} - \boldsymbol{A}^\top \boldsymbol{\lambda} = \mathbf{0} \\ \nabla_{\boldsymbol{\lambda}} l(\boldsymbol{x}, \boldsymbol{\lambda}) &= \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x} = \mathbf{0} \end{aligned}$$

Now we solve the following system for $[\boldsymbol{x}^*; \boldsymbol{\lambda}^*]$:

$$\nabla_{\boldsymbol{x}} L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \boldsymbol{Q}\boldsymbol{x}^* - \boldsymbol{A}^\top \boldsymbol{\lambda}^* = \boldsymbol{0}$$
$$\nabla_{\boldsymbol{\lambda}} L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}^* = \boldsymbol{0}$$

The first equation implies $\boldsymbol{x}^* = \boldsymbol{Q}^{-1}\boldsymbol{A}^\top \boldsymbol{\lambda}^*$.

Plugging this into the second equation and solve for $\boldsymbol{\lambda}^*$ to get

$$\boldsymbol{\lambda}^* = (\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^\top)^{-1}\boldsymbol{b}$$

Hence the solution is

$$\boldsymbol{x}^* = \boldsymbol{Q}^{-1}\boldsymbol{A}^\top \boldsymbol{\lambda}^* = \boldsymbol{Q}^{-1}\boldsymbol{A}^\top (\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^\top)^{-1}\boldsymbol{b}$$

Example. Consider the problem of finding the solution of minimal norm to the linear system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$. That is

$$\begin{aligned} \text{minimize} \quad & \|\boldsymbol{x}\| \\ \text{subject to} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{aligned}$$

Solution. The problem is equivalent to

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}\|\boldsymbol{x}\|^2 = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{x} \\ \text{subject to} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{aligned}$$

which is the problem above with $\boldsymbol{Q} = \boldsymbol{I}$. Hence the solution is

$$\boldsymbol{x}^* = \boldsymbol{A}^\top (\boldsymbol{A}\boldsymbol{A}^\top)^{-1} \boldsymbol{b}$$

Example. Consider a discrete dynamical system

$$x_k = ax_{k-1} + bu_k$$

with given initial x_0, where $k = 1, \dots, N$ stand for the time point. Here x_k is the "state" and u_k is the "control".

Suppose we want to minimize the state and control at all points, then we can formulate the problem as

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2} \sum_{k=1}^{N} (qx_k^2 + ru_k^2) \\ \text{subject to} \quad & x_k = ax_{k-1} + bu_k, \quad k = 1, \dots, N. \end{aligned}$$

This is an example of *linear quadratic regulator* (LQR) in the optimal control theory.

To solve this, we let $\boldsymbol{z} = [x_1; \dots; x_N; u_1; \dots; u_N] \in \mathbb{R}^{2N}$,

$$\boldsymbol{Q} = \begin{bmatrix} q\boldsymbol{I}_N & \boldsymbol{0} \\ \boldsymbol{0} & r\boldsymbol{I}_N \end{bmatrix} \in \mathbb{R}^{(2N)\times(2N)}$$

and

$$\boldsymbol{A} = \begin{bmatrix} 1 & & \cdots & 0 & -b & & \cdots & 0 \\ -a & 1 & & \vdots & & -b & & \vdots \\ & \ddots & \ddots & & \vdots & & \ddots & \\ 0 & & -a & 1 & 0 & \cdots & & -b \end{bmatrix} \quad \boldsymbol{b} = \begin{bmatrix} ax_0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Then the problem can be written as

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}\boldsymbol{z}^\top \boldsymbol{Q}\boldsymbol{z} \\ \text{subject to} \quad & \boldsymbol{A}\boldsymbol{z} = \boldsymbol{b} \end{aligned}$$

and the solution is

$$\boldsymbol{z}^* = [\boldsymbol{x}^*; \boldsymbol{u}^*] = \boldsymbol{Q}^{-1}\boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^\top)^{-1}\boldsymbol{b}$$

Example [Credit card holder's dilemma]. Suppose we have a credit card debt \$10,000 which has a monthly interest rate of 2%. Now we want to make monthly payment for 10 months to minimize the balance as well as the amount of monthly payments.

Let x_k be the balance and u_k be the payment in month k. Then the problem can be written as

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2} \sum_{k=1}^{10} (q x_k^2 + r u_k^2) \\ \text{subject to} \quad & x_k = 1.02\, x_{k-1} - u_k, \quad k = 1, \dots, 10, \quad x_0 = 10000. \end{aligned}$$

The more anxious we are to reduce our debt, the larger the value of q relative to r. On the other hand, the more reluctant we are to make payments, the larger the value of r relative to q.

Here are two instances with different choices of q and r:

$q = 1, r = 10$:

k	Balance x_k	Payment u_k
1	7326.60	2873.40
2	5374.36	2098.77
3	3951.13	1530.72
4	2916.82	1113.34
5	2169.61	805.54
6	1635.97	577.04
7	1263.35	405.34
8	1015.08	273.53
9	866.73	168.65
10	803.70	80.37

$q = 1, r = 300$:

k	Balance x_k	Payment u_k
1	9844.66	355.34
2	9725.36	316.20
3	9641.65	278.22
4	9593.23	241.25
5	9579.92	205.17
6	9601.68	169.84
7	9658.58	135.13
8	9750.83	100.92
9	9878.78	67.08
10	10042.87	33.48

We now focus on constrained optimization problems with both equality and inequality constraints:

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

and the feasible set is $\Omega = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{g}(\boldsymbol{x}) \le \mathbf{0},\ \boldsymbol{h}(\boldsymbol{x}) = \mathbf{0}\}$.

Note that $\boldsymbol{g}(\boldsymbol{x}) = [g_1(\boldsymbol{x}); \dots; g_p(\boldsymbol{x})] \in \mathbb{R}^p$ and $\boldsymbol{h}(\boldsymbol{x}) = [h_1(\boldsymbol{x}); \dots; h_m(\boldsymbol{x})] \in \mathbb{R}^m$.

Definition. We call the inequality constraint g_j **active** at $\boldsymbol{x} \in \Omega$ if $g_j(\boldsymbol{x}) = 0$ and **inactive** if $g_j(\boldsymbol{x}) < 0$.

Definition. Denote $J(\boldsymbol{x})$ the index set of active constraints at $\boldsymbol{x}$:

$$J(\boldsymbol{x}) = \{j : g_j(\boldsymbol{x}) = 0\}.$$

Also denote $J^c(\boldsymbol{x}) = \{1, \dots, p\} \setminus J(\boldsymbol{x})$ as its complement.

Definition. We call x a **regular point** in Ω if

$$\nabla h_i(\boldsymbol{x}), \quad \nabla g_j(\boldsymbol{x}), \quad 1 \le i \le m, \quad j \in J(\boldsymbol{x})$$

are linearly independent (total of $m + |J(\boldsymbol{x})|$ vectors in $\mathbb{R}^n$).

Now we consider the first order necessary condition (FONC) for the optimization problem with both equality and inequality constraints:

Theorem [Karush-Kahn-Tucker (KKT)]. Suppose $f, \boldsymbol{g}, \boldsymbol{h} \in C^1$, $\boldsymbol{x}^*$ is a regular point and local minimizer of f, then $\exists\ \boldsymbol{\lambda}^* \in \mathbb{R}^m,\ \boldsymbol{\mu}^* \in \mathbb{R}^p$ such that

$$
\begin{aligned}
\nabla f(\boldsymbol{x}^*)^\top + \boldsymbol{\lambda}^{*\top} D\boldsymbol{h}(\boldsymbol{x}^*) + \boldsymbol{\mu}^{*\top} D\boldsymbol{g}(\boldsymbol{x}^*) &= \mathbf{0}^\top \\
\boldsymbol{h}(\boldsymbol{x}^*) &= \mathbf{0} \\
\boldsymbol{g}(\boldsymbol{x}^*) &\le \mathbf{0} \\
\boldsymbol{\mu}^* &\ge \mathbf{0} \\
\boldsymbol{\mu}^{*\top} \boldsymbol{g}(\boldsymbol{x}^*) &= 0
\end{aligned}
$$

Remarks.

- Define **Lagrange function**:

$$l(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\boldsymbol{x}) + \boldsymbol{\lambda}^\top \boldsymbol{h}(\boldsymbol{x}) + \boldsymbol{\mu}^\top \boldsymbol{g}(\boldsymbol{x})$$

  then the first KKT condition is just $\nabla_{\boldsymbol{x}} l(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \boldsymbol{0}$.

- The second and third KKT conditions are just the constraints.

- $\boldsymbol{\lambda}$ is the **Lagrange multiplier** and $\boldsymbol{\mu}$ is the **KKT multiplier**.

- Since $\boldsymbol{\mu}^* \geq \boldsymbol{0}$ and $\boldsymbol{g}(\boldsymbol{x}^*) \leq \boldsymbol{0}$, the last KKT condition implies $\mu_j^* g_j(\boldsymbol{x}^*) = 0$ for all $j = 1, \dots, p$. Namely $g_j(\boldsymbol{x}^*) < 0$ implies $\mu_j^* = 0$. Hence

$$\mu_j^* = 0, \quad \forall\, j \neq J(\boldsymbol{x}^*).$$

Proof (KKT Theorem). We first just let $\mu_j^* = 0$ for all $j \in J^c(\boldsymbol{x}^*)$.

Since g_j is not active at $\boldsymbol{x}^*$ for $j \in J^c(\boldsymbol{x}^*)$, it's not active in a neighbor of $\boldsymbol{x}^*$ either. Hence $\boldsymbol{x}^*$ is a regular point and local minimizer in Ω implies that $\boldsymbol{x}^*$ is a regular point and local minimizer in

$$\Omega' := \{\boldsymbol{x} \in \Omega : \boldsymbol{h}(\boldsymbol{x}) = 0, g_j(\boldsymbol{x}) = 0, j \in J(\boldsymbol{x}^*)\}$$

Note Ω' only contains equality constraints, hence the Lagrange theorem for equality constrained problems applies, i.e., $\exists \boldsymbol{\lambda}^*, \mu_j^*$ for $j \in J(\boldsymbol{x}^*)$ such that

$$\nabla f(\boldsymbol{x}^*) + D\boldsymbol{h}(\boldsymbol{x}^*)^\top \boldsymbol{\lambda}^* + D\boldsymbol{g}(\boldsymbol{x}^*)^\top \boldsymbol{\mu}^* = \mathbf{0}$$

where $\boldsymbol{\mu}^* = [\mu_1^*; \dots; \mu_p^*]$. We only need to show $\mu_j^* \geq 0$ for all $j \in J(\boldsymbol{x}^*)$.

Proof (cont.) If $\mu_j^* < 0$ for some $j \in J(\boldsymbol{x}^*)$, then define

$$\widehat{T}(\boldsymbol{x}^*) := \{\boldsymbol{y} \in \mathbb{R}^n : D\boldsymbol{h}(\boldsymbol{x}^*)\boldsymbol{y} = \boldsymbol{0}, \nabla g_{j'}(\boldsymbol{x}^*)^\top \boldsymbol{y} = 0, j \in J(\boldsymbol{x}^*), j' \neq j\}.$$

We claim that $\exists\, \boldsymbol{y} \in \widehat{T}(\boldsymbol{x}^*)$ such that $\nabla g_j(\boldsymbol{x}^*)^\top \boldsymbol{y} \neq 0$: otherwise $\nabla g_j(\boldsymbol{x}^*)$ can be spanned by $\{\nabla h_i(\boldsymbol{x}^*), \nabla g_{j'}(\boldsymbol{x}^*) : 1 \leq i \leq m, j' \in J(\boldsymbol{x}^*), j' \neq j\}$, which contradicts to that $\nabla h_i(\boldsymbol{x}^*), \nabla g_j(\boldsymbol{x}^*)$ are linearly independent (since $\boldsymbol{x}^*$ is regular). We choose $\boldsymbol{y}$ (or $-\boldsymbol{y}$) so that $\nabla g_j(\boldsymbol{x}^*)^\top \boldsymbol{y} < 0$.

Now left-multiply $\boldsymbol{y}^\top$ to both sides of $\nabla f(\boldsymbol{x}^*) + D\boldsymbol{h}(\boldsymbol{x}^*)^\top \boldsymbol{\lambda}^* + D\boldsymbol{g}(\boldsymbol{x}^*)^\top \boldsymbol{\mu}^* = \boldsymbol{0}$, we get (since $\mu_j^* < 0$ and $\nabla g_j(\boldsymbol{x}^*)^\top \boldsymbol{y} < 0$):

$$0 = \boldsymbol{y}^\top \nabla f(\boldsymbol{x}^*) + \mu_j^* \boldsymbol{y}^\top \nabla g_j(\boldsymbol{x}^*) > \boldsymbol{y}^\top \nabla f(\boldsymbol{x}^*)$$

Therefore there exists a curve $\boldsymbol{x}(t) : (a, b) \to \Omega$ such that $\boldsymbol{x}(t^*) = \boldsymbol{x}^*$ and $\boldsymbol{x}'(t^*) = \boldsymbol{y}$ for $t^* \in (a, b)$.

Proof (cont.) Moreover, define $\phi(t) := f(\boldsymbol{x}(t))$, then

$$\phi'(t^*) = \nabla f(\boldsymbol{x}(t^*))^\top \boldsymbol{x}'(t^*) = \nabla f(\boldsymbol{x}^*)^\top \boldsymbol{y} < 0$$

Also, define $\psi(t) = g_j(\boldsymbol{x}(t))$, then

$$\psi'(t^*) = \nabla g_j(\boldsymbol{x}(t^*))^\top \boldsymbol{x}'(t^*) = \nabla g_j(\boldsymbol{x}^*)^\top \boldsymbol{y} < 0$$

These mean that $\exists\, \epsilon > 0$ such that during $[t^*, t^* + \epsilon] \subset (a, b)$, $f(\boldsymbol{x}(t))$ and $g_j(\boldsymbol{x}(t))$ can both decrease further, so $\boldsymbol{x}(t) \in \Omega$ and $f(\boldsymbol{x}(t)) < f(\boldsymbol{x}^*)$ for $t \in (t^*, t^* + \epsilon]$. This contradicts to that $\boldsymbol{x}^*$ is a local minimizer on Ω. Hence $\mu_j^* \geq 0$ for all $j \in J(\boldsymbol{x}^*)$.

Example. Consider the problem

$$\begin{aligned} \text{minimize} \quad & f(x_1, x_2) = x_1^2 + x_2^2 + x_1 x_2 - 3x_1 \\ \text{subject to} \quad & x_1, x_2 \geq 0 \end{aligned}$$

The Lagrange function is

$$l(\boldsymbol{x}, \boldsymbol{\mu}) = x_1^2 + x_2^2 + x_1 x_2 - 3x_1 - x_1 \mu_1 - x_2 \mu_2$$

The KKT condition is

$$\begin{aligned} 2x_1 + x_2 - 3 - \mu_1 &= 0 \\ x_1 + 2x_2 - \mu_2 &= 0 \\ x_1, x_2, \mu_1, \mu_2 &\geq 0 \\ \mu_1 x_1 + \mu_2 x_2 &= 0 \end{aligned}$$

Solving this yields

$$x_1^* = \mu_2^* = \frac{3}{2}, \quad x_2^* = \mu_1^* = 0.$$

Similar as the proof of FONC, we can show SONC.

Theorem [Second order necessary condition (SONC)]. Suppose $f, \boldsymbol{g}, \boldsymbol{h} \in C^2$. If $\boldsymbol{x}^*$ is a regular point and local minimizer, then $\exists\, \boldsymbol{\lambda}^* \in \mathbb{R}^m, \boldsymbol{\mu}^* \in \mathbb{R}^p_+$ such that

- The KKT condition for $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ holds;
- For all $\boldsymbol{y} \in T(\boldsymbol{x}^*)$, there is

$$\boldsymbol{y}^\top \nabla^2_{\boldsymbol{x}} l(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)\boldsymbol{y} \geq 0$$

where

$$T(\boldsymbol{x}^*) = \{\boldsymbol{y} \in \mathbb{R}^n : D\boldsymbol{h}(\boldsymbol{x}^*)\boldsymbol{y} = \boldsymbol{0}, \nabla g_j(\boldsymbol{x}^*)^\top \boldsymbol{y} = 0, \forall\, j \in J(\boldsymbol{x}^*)\}$$

Proof. The first part follows from the KKT theorem. The second part is due to the fact that $\boldsymbol{x}^*$ being a local minimizer of f over Ω implies that it is a local minimizer over Ω'.

Theorem [Second order sufficient condition (SOSC)]. Suppose $f, \boldsymbol{g}, \boldsymbol{h} \in C^2$. If $\exists\, \boldsymbol{\lambda}^* \in \mathbb{R}^m, \boldsymbol{\mu}^* \in \mathbb{R}^p$ such that

- The KKT condition of $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ holds;
- For all nonzero $\boldsymbol{y} \in \tilde{T}(\boldsymbol{x}^*, \boldsymbol{\mu}^*)$, there is

$$\boldsymbol{y}^\top \nabla_{\boldsymbol{x}}^2 l(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)\boldsymbol{y} > 0$$

  where

$$\tilde{T}(\boldsymbol{x}^*, \boldsymbol{\mu}^*) := \{\boldsymbol{y} \in \mathbb{R}^n : D\boldsymbol{h}(\boldsymbol{x})\boldsymbol{y} = \boldsymbol{0}, \nabla g_j(\boldsymbol{x}^*)\boldsymbol{y} = 0, j \in \tilde{J}(\boldsymbol{x}^*, \boldsymbol{\mu}^*)\}$$

  and

$$\tilde{J}(\boldsymbol{x}^*, \boldsymbol{\mu}^*) := \{j \in J(\boldsymbol{x}^*) : \mu_j^* > 0\}$$

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

Remark. We omit the proof here. Note that $T(\boldsymbol{x}^*) \subset \tilde{T}(\boldsymbol{x}^*, \boldsymbol{\mu}^*)$.

Example. Consider the following constrained problem:

$$\begin{aligned} \text{minimize} \quad & x_1 x_2^2 \\ \text{subject to} \quad & x_1 = x_2 \\ & x_1 \geq 0 \end{aligned}$$

Solution. Here $f(\boldsymbol{x}) = x_1 x_2^2$, $h(\boldsymbol{x}) = x_1 - x_2$, and $g(\boldsymbol{x}) = -x_1$. The Lagrange function is

$$l(\boldsymbol{x}, \boldsymbol{\lambda}, \mu) = x_1 x_2^2 + \lambda(x_1 - x_2) - \mu x_1$$

Then we obtain the KKT conditions:

$$\begin{aligned} \partial_{x_1} l(\boldsymbol{x}, \lambda, \mu) &= x_2^2 + \lambda - \mu = 0 \\ \partial_{x_2} l(\boldsymbol{x}, \lambda, \mu) &= 2x_1 x_2 - \lambda = 0 \\ \partial_{\lambda} l(\boldsymbol{x}, \lambda, \mu) &= x_1 - x_2 = 0 \\ x_1 &\geq 0 \\ \mu &\geq 0 \\ \mu x_1 &= 0 \end{aligned}$$

Solution (cont.) If $x_1^* = x_2^* = 0$, then $\lambda^* = \mu^* = 0$. If $x_1^* = x_2^* > 0$, then $\mu^* = 0$ but we cannot find any valid λ^*. So only the point $[\boldsymbol{x}^*, \lambda^*, \mu^*] = [0, 0, 0, 0]$ satisfies the KKT conditions.

Since $\mu^* = 0$, we have

$$\tilde{T}(\boldsymbol{x}^*, \mu^*) = \mathcal{N}(\nabla h(\boldsymbol{x}^*)) = \mathcal{N}([1, -1]) = \{t[1, 1] : t \in \mathbb{R}\}$$

On the other hand

$$\nabla_{\boldsymbol{x}}^2 l(\boldsymbol{x}^*, \lambda^*, \mu^*) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

so $\boldsymbol{y}^\top (\nabla^2 l(\boldsymbol{x}^*, \lambda^*, \mu^*))\boldsymbol{y} = 0$ for all $\boldsymbol{y} \in \tilde{T}(\boldsymbol{x}^*, \mu)$ but not strictly larger than 0. Hence SOSC does not hold. But in fact $\boldsymbol{x}^* = [0, 0]$ is the local minimum (actually also global).

Example. Consider the following constrained problem:

$$\begin{aligned} \text{minimize} \quad & x_1 + 4x_2^2 \\ \text{subject to} \quad & x_1^2 + 2x_2^2 \ge 4 \end{aligned}$$

Solution. Here $f(\boldsymbol{x}) = x_1^2 + 4x_2^2$, $g(\boldsymbol{x}) = -(x_1^2 + 2x_2^2 - 4)$. The Lagrange function is

$$l(\boldsymbol{x}, \mu) = x_1^2 + 4x_2^2 - \mu(x_1^2 + 2x_2^2 - 4).$$

Then we obtain the KKT conditions:

$$\begin{aligned} \partial_{x_1} l(\boldsymbol{x}, \mu) = 2x_1 - 2\mu x_1 &= 0 \\ \partial_{x_2} l(\boldsymbol{x}, \mu) = 8x_2 - 4\mu x_2 &= 0 \\ x_1^2 + 2x_2^2 &\ge 4 \\ \mu &\ge 0 \\ -\mu(x_1^2 + 2x_2^2 - 4) &= 0 \end{aligned}$$

Solution (cont.)

- If $\mu^* = 0$, then $x_1^* = x_2^* = 0$ which violates $g(\boldsymbol{x}) \le 0$.
- If $\mu^* = 1$ then $[x_1^*, x_2^*] = \pm[2, 0]$.
- If $\mu^* = 2$ then $[x_1^*, x_2^*] = \pm[0, \sqrt{2}]$.
- If $\mu^* > 0$ but $\mu \neq 1, 2$, then $x_1^* = x_2^* = 0$ which again violates $g(\boldsymbol{x}) \le 0$.

Hence the following 4 points satisfy the KKT conditions:

$$\begin{bmatrix} x_1^* \\ x_2^* \\ \mu^* \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}, \quad \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ \sqrt{2} \\ 2 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ -\sqrt{2} \\ 2 \end{bmatrix}$$

Solution (cont.)

For $\mu^* = 1$, we have

$$\nabla^2_{\boldsymbol{x}} l([\pm 2, 0, 1]) = \begin{bmatrix} 0 & 0 \\ 0 & 4 \end{bmatrix}, \quad \nabla g([\pm 2, 0]) = \begin{bmatrix} \mp 4 \\ 0 \end{bmatrix}$$

which implies

$$\tilde{T}(\boldsymbol{x}^*, \mu^*) = T(\boldsymbol{x}^*) = \{t[0, 1] : t \in \mathbb{R}\}$$

Hence

$$\boldsymbol{y}^\top \nabla^2_{\boldsymbol{x}} l([x_1^*, x_2^*, \mu^*])\boldsymbol{y} = 4t^2 > 0$$

for all $\boldsymbol{y} \in \tilde{T}(\boldsymbol{x}^*, \mu^*) \setminus \{\boldsymbol{0}\}$.

So $[x_1^*, x_2^*] = [\pm 2, 0]$ satisfy SOSC and are strict local minimizers.

Solution (cont.)

For $\mu^* = 2$, we have

$$\nabla^2_{\boldsymbol{x}} l([0, \pm\sqrt{2}, 2]) = \begin{bmatrix} -2 & 0 \\ 0 & 0 \end{bmatrix}, \quad \nabla g([0, \pm\sqrt{2}]) = \begin{bmatrix} 0 \\ \mp 4\sqrt{2} \end{bmatrix}$$

which implies

$$\tilde{T}(\boldsymbol{x}^*, \mu^*) = T(\boldsymbol{x}^*) = \{t[1, 0] : t \in \mathbb{R}\}$$

Hence

$$\boldsymbol{y}^\top \nabla^2_{\boldsymbol{x}} l([x_1^*, x_2^*, \mu^*])\boldsymbol{y} = -4t^2 < 0$$

for all $\boldsymbol{y} \in \tilde{T}(\boldsymbol{x}^*, \mu^*) \setminus \{\boldsymbol{0}\}$.

So $[x_1^*, x_2^*] = [0, \pm\sqrt{2}]$ do not satisfy SOSC but are strict local maximizers.