ROYAL INSTITUTE
OF TECHNOLOGY

Lecture: NLP with equality constraints

1. Nonlinear Programming with equality constraints.
2. Optimality conditions

General nonlinear problems under equality constraints

The general problem is

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & h_i(\mathbf{x}) = 0, \; i = 1, \ldots, m \end{aligned} \tag{1}$$

The feasible region $\mathcal{F} = \{\mathbf{x} \in \mathbf{R}^n : h_i(\mathbf{x}) = 0, \; i = 1, \ldots, m\}$ is in general not convex.

We will start by considering a simpler convex case, namely, the case when the functions h_i are affine, *i.e.*, $h_i(\mathbf{x}) = \mathbf{a}_i^\mathsf{T}\mathbf{x} + b_i$. We assume that $n > m$.

The feasible region $\mathcal{F} = \{\mathbf{x} \in \mathbf{R}^n : \mathbf{A}\mathbf{x} = \mathbf{b}, \}$ is now convex, and we assume that the rows of $\mathbf{A}$ are linearly independent.

NLP with linear equality constraints

Use a **nullspace method** to solve

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} = b, \end{aligned} \tag{2}$$

If $\bar{\mathbf{x}}$ is an arbitrary feasible point, then any $\mathbf{x} \in \mathcal{F}$ can be written $\mathbf{x} = \bar{\mathbf{x}} + Zv$ where the columns of Z span the nullspace of $\mathbf{A}$.

(2) is equivalent to minimize $\varphi(v) = f(\bar{\mathbf{x}} + Zv)$ s.t. $v \in \mathbf{R}^{n-m}$.

The first order optimality condition is

$$\nabla_v f(v) = \nabla_x f(\bar{\mathbf{x}} + Zv)\nabla_v(\bar{\mathbf{x}} + Zv) = \nabla_x f(\bar{\mathbf{x}} + Zv)Z = 0$$

NLP with linear equality constraints

A **Lagrange approach**

Know: $\nabla f(x_*)^T \in \mathbf{R}^n = \mathcal{N}(\mathbf{A}) \oplus \mathcal{R}(\mathbf{A}^T)$,
so $\nabla f(x_*)^T = Zv_* + \mathbf{A}^T\lambda_*$ for some vectors v_* and λ_*.

If x_* is a local minimum, we know $Z^T\nabla f(x_*)^T = 0$,
i.e. $Z^T(Zv_* + \mathbf{A}^T\lambda_*) = Z^TZv_* + \underbrace{Z^T\mathbf{A}^T}_{=0}\lambda_* = 0$.

So $Z^TZv_* = 0$, hence $Zv_* = 0$ and then $\nabla f(x_*)^T = \mathbf{A}^T\lambda_*$ must hold at a local minimum for (2).

NLP under general equality constraints

Consider again the general problem

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & h_i(\mathbf{x}) = 0, \; i = 1, \ldots, m \end{aligned} \tag{3}$$

For the linear case we assumed that the rows of $\mathbf{A}$ where linearly independent, now we need the following technical assumption:

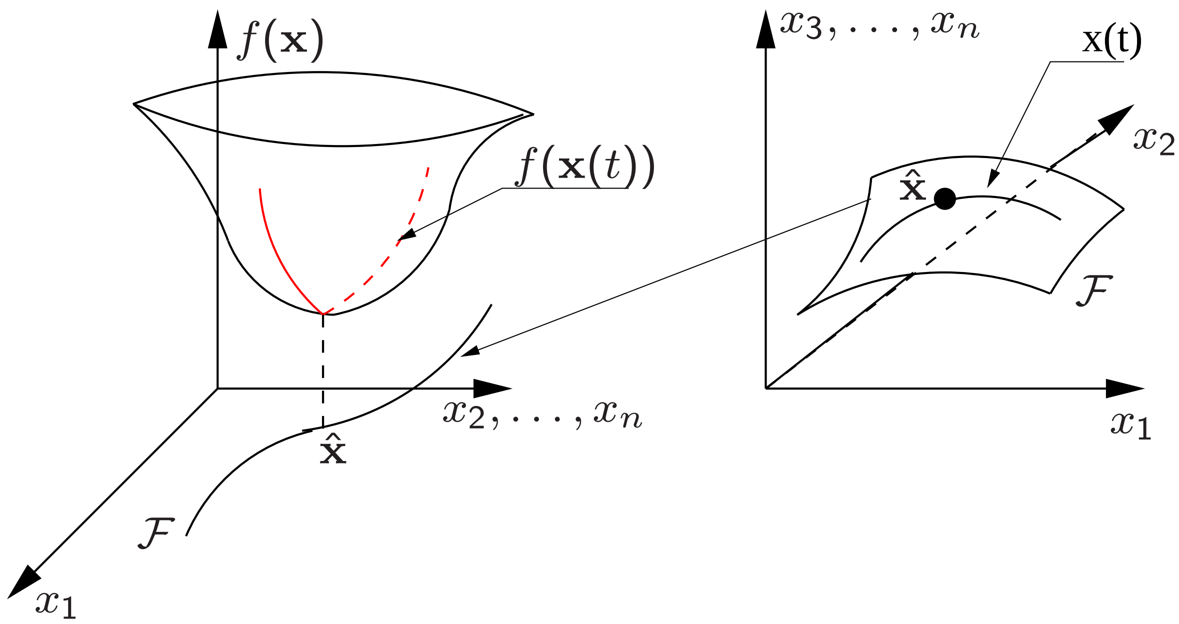
Definition 1. *A feasible solution $\mathbf{x} \in \mathcal{F}$ is a regular point to (1) if $\nabla h_i(\mathbf{x})$, $i = 1, \ldots, m$ are linearly independent.*

Theorem 1 (Lagrange's optimality conditions)**.** *Assume that* $\hat{\mathbf{x}} \in \mathcal{F}$ *is a regular point and a local optimal solution to (1). Then there exists* $\hat{\mathbf{u}} \in \mathbf{R}^m$ *such that*

(1) $\nabla f(\hat{\mathbf{x}}) + \sum_{i=1}^{m} \hat{u}_i \nabla h_i(\hat{\mathbf{x}}) = \mathbf{0}^\mathsf{T}$,

(2) $h_i(\hat{\mathbf{x}}) = 0$, $i = 1, \ldots, m$.

Proof idea: *Let* $\mathbf{x}(t)$ *be an arbitrary parameterized curve in the feasible set* $\mathcal{F} = \{\mathbf{x} \in \mathbf{R}^n : h_i(\mathbf{x}) = 0,\ i = 1, \ldots, m\}$ *such that* $\mathbf{x}(0) = \hat{\mathbf{x}}$*. The figure on the next page illustrates how this curve is mapped on a curve* $f(\mathbf{x}(t))$ *on the range space of the objective function. The feasible set* $\mathcal{F}$ *is in general of higher dimension than one, which is illustrated in the right figure.*



Since $\mathbf{x}(0) = \hat{\mathbf{x}}$ is a local optimal solution it holds that

$$\frac{d}{dt} f(\mathbf{x}(t))|_{t=0} = \nabla f(\hat{\mathbf{x}}) \cdot \mathbf{x}'(0) = 0$$

Furthermore, $\mathbf{x}(t) \in \mathcal{F}$, which leads to

$$h_i(\mathbf{x}(t)) = 0, \quad i = 1, \ldots, m, \quad \forall t \in (-\epsilon, \epsilon)$$

for some $\epsilon > 0$.

This means that

$$\frac{d}{dt} h_i(\mathbf{x}(t))|_{t=0} = \nabla h_i(\hat{\mathbf{x}}) \cdot \mathbf{x}'(0) = 0, \qquad i = 1, \ldots, m$$

which in turn leads to $\mathbf{x}'(0) \in \mathcal{N}(\mathbf{A})$, where

$$\mathbf{A} = \begin{bmatrix} \nabla h_1(\hat{\mathbf{x}}) \\ \vdots \\ \nabla h_m(\hat{\mathbf{x}}) \end{bmatrix}$$

Conversely, the implicit function theorem can be used to show that if $\mathbf{p} \in \mathcal{N}(\mathbf{A})$, then there exists a parameterized curve $\mathbf{x}(t) \in \mathcal{F}$ with $\mathbf{x}(0) = \hat{\mathbf{x}}$ and $\mathbf{x}'(0) = \mathbf{p}$.

Alltogether, the above argument shows that

$$\begin{aligned}
& \nabla f(\hat{\mathbf{x}})\mathbf{p} = 0, \qquad \forall \mathbf{p} \in \mathcal{N}(\mathbf{A}) \\
\Leftrightarrow \quad & \nabla f(\hat{\mathbf{x}})^{\mathsf{T}} \in \mathcal{N}(\mathbf{A})^{\perp} = \mathcal{R}(\mathbf{A}^{\mathsf{T}}) \\
\Leftrightarrow \quad & \nabla f(\hat{\mathbf{x}})^{\mathsf{T}} = \mathbf{A}^{\mathsf{T}}\hat{\mathbf{v}},
\end{aligned}$$

for some $\hat{\mathbf{v}} \in \mathbf{R}^m$. If we let $\hat{\mathbf{u}} = -\hat{\mathbf{v}} \in \mathbf{R}^m$ the last expression can be written

$$\nabla f(\hat{\mathbf{x}}) + \sum_{i=1}^{m} \hat{u}_i \nabla h_i(\hat{\mathbf{x}}) = \mathbf{0}^{\mathsf{T}}$$

which was to be proven.

Example

Consider

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & h(\mathbf{x}) = 0, \end{aligned} \tag{4}$$

where $f(x) = x_1 x_2 - \log |x_1|$ and $h(x) = x_1 - x_2 - 2$.

The constraint is linear and can be written $\mathbf{A}\mathbf{x} = \mathbf{b}$, where

$$\mathbf{A} = \begin{bmatrix} 1 & -1 \end{bmatrix}, \quad \mathbf{b} = 2, \text{ and } Z = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

the matrix Z spans the nullspace of A

Then $\nabla f(\mathbf{x}) = \begin{bmatrix} x_2 - 1/x_1 & x_1 \end{bmatrix}$

We want to determine optimality conditions and find all points satisfying them.

Example - Nullspace method

The reduced gradient is given by

$$Z^T \nabla f(\mathbf{x})^T = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} x_2 - 1/x_1 \\ x_1 \end{bmatrix} = x_2 - 1/x_1 + x_1.$$

Setting it equal to zero and using that $x_2 = x_1 - 2$, we get $x_1^2 - x_1 - 1/2 = 0$, with solutions

$$\mathbf{x}^{(1)} = (\frac{1+\sqrt{3}}{2}, \frac{-3+\sqrt{3}}{2}), \quad \mathbf{x}^{(2)} = (\frac{1-\sqrt{3}}{2}, \frac{-3-\sqrt{3}}{2}).$$

We get

$$f(\mathbf{x}^{(1)}) > f(\mathbf{x}^{(2)}),$$

so $\mathbf{x}^{(2)}$ is the best stationary point.

Example - Lagrange method

We check that $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$ satisfy the conditions $\nabla f(\mathbf{x}^{(k)})^T + \lambda_k \nabla h(\mathbf{x}^{(k)})^T = 0$ for some λ_k when $k = 1, 2$.

$$\nabla f(\mathbf{x}^{(1)})^T + \lambda_1 \nabla h(\mathbf{x}^{(1)})^T = \begin{bmatrix} -\frac{1+\sqrt{3}}{2} \\ \frac{1+\sqrt{3}}{2} \end{bmatrix} + \lambda_1 \underbrace{\begin{bmatrix} 1 \\ -1 \end{bmatrix}}_{=\mathbf{A}^T} = 0$$

which is satisfied for $\lambda_1 = \frac{1+\sqrt{3}}{2}$.

$$\nabla f(\mathbf{x}^{(2)})^T + \lambda_2 \nabla h(\mathbf{x}^{(2)})^T = \begin{bmatrix} -\frac{1-\sqrt{3}}{2} \\ \frac{1-\sqrt{3}}{2} \end{bmatrix} + \lambda_1 \underbrace{\begin{bmatrix} 1 \\ -1 \end{bmatrix}}_{=\mathbf{A}^T} = 0$$

which is satisfied for $\lambda_2 = \frac{1-\sqrt{3}}{2}$.

Example - Graphical illustration

The function f is depicted below, in $\mathbf{R}^2$ (left), for feasible points (right).