Lecture: NLP with equality constraints

1. Nonlinear Programming with equality constraints.

2. Optimality conditions
General nonlinear problems under equality constraints

The general problem is

\[
\begin{align*}
\text{minimize} & \quad f(x) \\
\text{s.t.} & \quad h_i(x) = 0, \ i = 1, \ldots, m
\end{align*}
\]  

(1)

The feasible region \( F = \{ x \in \mathbb{R}^n : h_i(x) = 0, \ i = 1, \ldots, m \} \) 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., \( h_i(x) = a_i^T x + b_i \). We assume that \( n > m \).

The feasible region \( F = \{ x \in \mathbb{R}^n : Ax = b, \} \) is now convex, and we assume that the rows of \( A \) are linearly independent.
NLP with linear equality constraints

Use a nullspace method to solve

\[
\begin{align*}
\text{minimize} & \quad f(x) \\
\text{s.t.} & \quad Ax = b,
\end{align*}
\]  

(2)

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

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

The first order optimality condition is

\[
\nabla_v f(v) = \nabla_x f(\bar{x} + Zv) \nabla_v (\bar{x} + Zv) = \nabla_x f(\bar{x} + Zv) Z = 0
\]
**A Lagrange approach**

Know: \( \nabla f(x^\ast)^T \in \mathbb{R}^n = \mathcal{N}(A) \oplus \mathcal{R}(A^T) \),

so \( \nabla f(x^\ast)^T = Zv^\ast + A^T\lambda^\ast \) for some vectors \( v^\ast \) and \( \lambda^\ast \).

If \( x^\ast \) is a local minimum, we know \( Z^T \nabla f(x^\ast)^T = 0 \),
i.e. \( Z^T(Zv^\ast + A^T\lambda^\ast) = Z^TZv^\ast + \underbrace{Z^TA^T}_{=0} \lambda^\ast = 0 \).

So \( Z^TZv^\ast = 0 \), hence \( Zv^\ast = 0 \) and then \( \nabla f(x^\ast)^T = A^T\lambda^\ast \) must hold at a local minimum for (2).
Consider again the general problem

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

(3)

For the linear case we assumed that the rows of $A$ are linearly independent, now we need the following technical assumption:

**Definition 1.** A feasible solution $x \in \mathcal{F}$ is a regular point to (1) if $\nabla h_i(x), \quad i = 1, \ldots, m$ are linearly independent.
Theorem 1 (Lagrange’s optimality conditions). Assume that $\hat{x} \in \mathcal{F}$ is a regular point and a local optimal solution to (1). Then there exists $\hat{u} \in \mathbb{R}^m$ such that

\begin{align*}
(1) \quad \nabla f(\hat{x}) + \sum_{i=1}^{m} \hat{u}_i \nabla h_i(\hat{x}) &= 0^T, \\
(2) \quad h_i(\hat{x}) &= 0, \quad i = 1, \ldots, m.
\end{align*}

Proof idea: Let $x(t)$ be an arbitrary parameterized curve in the feasible set $\mathcal{F} = \{x \in \mathbb{R}^n : h_i(x) = 0, \quad i = 1, \ldots, m\}$ such that $x(0) = \hat{x}$. The figure on the next page illustrates how this curve is mapped on a curve $f(x(t))$ 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 $x(0) = \hat{x}$ is a local optimal solution it holds that

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

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

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

for some $\epsilon > 0$. 
This means that
\[
\frac{d}{dt} h_i(x(t)) \bigg|_{t=0} = \nabla h_i(\hat{x}) \cdot x'(0) = 0, \quad i = 1, \ldots, m
\]
which in turn leads to \(x'(0) \in \mathcal{N}(A)\), where

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

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

\[ \nabla f(\hat{x})p = 0, \quad \forall p \in \mathcal{N}(A) \]

\[ \iff \nabla f(\hat{x})^T \in \mathcal{N}(A)^\perp = \mathcal{R}(A^T) \]

\[ \iff \nabla f(\hat{x})^T = A^T \hat{v}, \]

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

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

which was to be proven.
Example

Consider

\[
\begin{align*}
\text{minimize} & \quad f(x) \\
\text{s.t.} & \quad h(x) = 0,
\end{align*}
\]

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 \( Ax = b \), where

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

the matrix \( Z \) spans the nullspace of \( A \).

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

We want to determine optimality conditions and find all points satisfying them.
Example - Nullspace method

The reduced gradient is given by

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

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

\[ x^{(1)} = \left( \frac{1 + \sqrt{3}}{2}, -3 + \sqrt{3} \right), \quad x^{(2)} = \left( \frac{1 - \sqrt{3}}{2}, -3 - \sqrt{3} \right). \]

We get

\[ f(x^{(1)}) > f(x^{(2)}), \]

so \( x^{(2)} \) is the best stationary point.
Example - Lagrange method

We check that $x^{(1)}$ and $x^{(2)}$ satisfy the conditions

$$\nabla f(x^{(k)})^T + \lambda_k \nabla h(x^{(k)})^T = 0$$

for some $\lambda_k$ when $k = 1, 2$.

\[
\nabla f(x^{(1)})^T + \lambda_1 \nabla h(x^{(1)})^T = \begin{bmatrix}
-\frac{1+\sqrt{3}}{2} \\
\frac{1+\sqrt{3}}{2}
\end{bmatrix} + \lambda_1 \begin{bmatrix} 1 \\ -1 \end{bmatrix} = 0
\]

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

\[
\nabla f(x^{(2)})^T + \lambda_2 \nabla h(x^{(2)})^T = \begin{bmatrix}
-\frac{1-\sqrt{3}}{2} \\
\frac{1-\sqrt{3}}{2}
\end{bmatrix} + \lambda_1 \begin{bmatrix} 1 \\ -1 \end{bmatrix} = 0
\]

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

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