Lagrange relaxation and KKT conditions

1. Lagrange relaxation
   - Global optimality conditions
   - KKT conditions for convex problems
   - Applications
2. KKT conditions for general nonlinear optimization problems.

Lagrange relaxation

We consider the optimization problem

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

where $f(\mathbf{x})$ and $g_i(\mathbf{x})$ are real valued functions.

If $\mathbf{g}(\mathbf{x}) = \begin{bmatrix} g_1(\mathbf{x}) & \ldots & g_m(\mathbf{x}) \end{bmatrix}^\mathsf{T}$ then (1) can be written

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{g}(\mathbf{x}) \le \mathbf{0}. \end{aligned} \tag{2}$$

The idéa behind Lagrange relaxation is to put non-negative prices $y_i \geq 0$, on the constraints and then add these to the objective function. This gives the (unconstrained) optimization problem:

$$\text{minimize} \quad f(\mathbf{x}) + \sum_{i=1}^{m} y_i g_i(\mathbf{x}) \tag{3}$$

which using $\mathbf{y} = \begin{bmatrix} y_1 & \dots & y_m \end{bmatrix}^\mathsf{T}$ can be written

$$\text{minimize} \quad f(\mathbf{x}) + \mathbf{y}^\mathsf{T}\mathbf{g}(\mathbf{x})$$

The "price" y_i is called a Lagrange multiplicator.

Definition 1. *The function $L : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}$ defined by $L(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}) + \mathbf{y}^\mathsf{T}\mathbf{g}(\mathbf{x})$ is called the Lagrange function to (2).*

Weak duality

Theorem 1 (Weak duality). *For an arbitrary* $\mathbf{y} \geq \mathbf{0}$ *it holds that* $\min_{\mathbf{x}} L(\mathbf{x}, \mathbf{y}) \leq f(\hat{\mathbf{x}})$, *where* $\hat{\mathbf{x}}$ *is an optimal solution to (2).*

Proof: Since $\hat{\mathbf{x}}$ is a feasible solution to (2) it holds that $\mathbf{g}(\hat{\mathbf{x}}) \leq \mathbf{0}$. We get

$$\min_{\mathbf{x}} L(\mathbf{x}, \mathbf{y}) \leq L(\hat{\mathbf{x}}, \mathbf{y}) = f(\hat{\mathbf{x}}) + \underbrace{\mathbf{y}^\mathsf{T}}_{\geq 0} \underbrace{\mathbf{g}(\hat{\mathbf{x}})}_{\leq 0} \leq f(\hat{\mathbf{x}}). \tag{4}$$

- minimizing the Lagrange function provides lower bounds to the optimization problem (2).
- By an appropriate choice of $\mathbf{y}$ a good approximation of the optimal solution to (2) is searched for. In practical algorithms one tries to solve $\max_{\mathbf{y}\geq\mathbf{0}} \min_{\mathbf{x}} L(\mathbf{x}, \mathbf{y})$. The next theorem gives conditions for the Lagrange multiplicator providing equality in (4).

Global optimality conditions

Theorem 2. *If* $(\hat{\mathbf{x}}, \hat{\mathbf{y}}) \in \mathbf{R}^n \times \mathbf{R}^m$ *satisfies the conditions*

(1) $L(\hat{\mathbf{x}}, \hat{\mathbf{y}}) = \min_x L(\mathbf{x}, \hat{\mathbf{y}})$,

(2) $\mathbf{g}(\hat{\mathbf{x}}) \leq \mathbf{0}$,

(3) $\hat{\mathbf{y}} \geq 0$,

(4) $\hat{\mathbf{y}}^{\mathsf{T}}\mathbf{g}(\hat{\mathbf{x}}) = 0$.

then $\hat{\mathbf{x}}$ *is an optimal solution to (2).*

Proof: *If* $\mathbf{x}$ *is an arbitrary feasible solution to (2) it holds that* $\mathbf{g}(\mathbf{x}) \leq \mathbf{0}$, *which shows that*

$$f(\mathbf{x}) \geq f(\mathbf{x}) + \hat{\mathbf{y}}^{\mathsf{T}}\mathbf{g}(\mathbf{x}) = L(\mathbf{x}, \hat{\mathbf{y}}) \geq L(\hat{\mathbf{x}}, \hat{\mathbf{y}}) = f(\hat{\mathbf{x}})$$

where the first inequality follows from (3) *and* $g(\mathbf{x}) \leq \mathbf{0}$, *the second inequality follows from* (1), *and the last one from* (4).

Convex optimization problems

- If the functions f and $g_1, \ldots, g_m$ are convex and continuously differentiable, then condition (1) in Theorem 2 is equivalent to the condition

$$\nabla f(\hat{\mathbf{x}}) + \sum_{i=1}^{m} \hat{y}_i \nabla g_i(\hat{\mathbf{x}}) = \mathbf{0}^{\mathsf{T}} \tag{5}$$

  This follows since $L(\mathbf{x}, \hat{\mathbf{y}})$ is convex when $\hat{\mathbf{y}} \geq 0$ and then it holds that $\hat{\mathbf{x}}$ is a minimum point for $L(\mathbf{x}, \hat{\mathbf{y}})$ if, and only if, $\nabla_{\mathbf{x}} L(\hat{\mathbf{x}}, \hat{\mathbf{y}}) = \mathbf{0}^{\mathsf{T}}$, *i.e.*, if and only if (5) is satisfied.

- The global optimality conditions in Theorem 2 are sufficient conditions for optimality, but in general not necessary. The next theorem shows that they are often also necessary conditions for convex optimization problems.

Definition 2. *The optimization problem (1) is a regular convex optimization problem if the functions f and $g_1, \ldots, g_m$ are convex and continuously differentiable and there exists a point $\mathbf{x}_0 \in \mathbf{R}^n$ such that $g_i(\mathbf{x}_0) < 0$, $i = 1, \ldots, m$.*

Theorem 3 (KKT for convex problems). *Assume that (1) is a regular convex problem. Then $\hat{\mathbf{x}}$ is a (global) optimal solution if, and only if, there exists a vector $\hat{\mathbf{y}} \in \mathbf{R}^m$ such that*

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

(2) $\mathbf{g}(\hat{\mathbf{x}}) \leq \mathbf{0}$,

(3) $\hat{\mathbf{y}} \geq \mathbf{0}$,

(4) $\hat{\mathbf{y}}^{\mathsf{T}} \mathbf{g}(\hat{\mathbf{x}}) = 0$.

Proof: *Sufficiency was shown previously. Necessity is shown in the book.*

The conditions (2) – (4) can be made more explicit. We have that

$$\hat{\mathbf{y}}^{\mathsf{T}}\mathbf{g}(\hat{\mathbf{x}}) = \sum_{i=1}^{m} \hat{y}_i g_i(\hat{\mathbf{x}}) = 0$$

Since $g_i(\hat{\mathbf{x}}) \leq 0$ and $\hat{y}_i \geq 0$ it follows that $\hat{y}_i g_i(\hat{\mathbf{x}}) = 0$, $i = 1, \ldots, m$. We then get the equivalent conditions

(2′) $g_i(\hat{\mathbf{x}}) \leq 0$, $i = 1, \ldots, m$,

(3′) $\hat{y}_i \geq 0$, $i = 1, \ldots, m$,

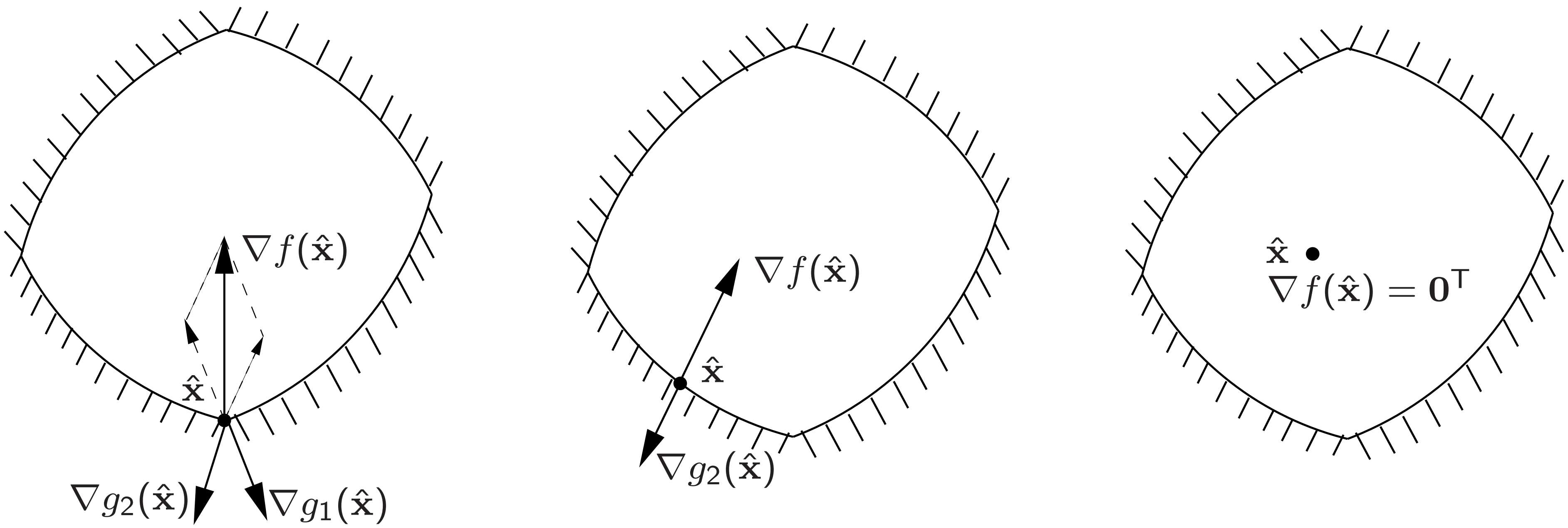
(4′) $\hat{y}_i \cdot g_i(\hat{\mathbf{x}}) = 0$, $i = 1, \ldots, m$.

Geometric interpretation

The complementarity condition (4′) implies that if $g_i(\hat{\mathbf{x}}) < 0$ then $y_i = 0$. Therefore, condition (1) can be written

$$\nabla f(\hat{\mathbf{x}}) = - \sum_{i:g_i(\hat{\mathbf{x}})=0} \hat{y}_i \nabla g_i(\hat{\mathbf{x}})$$

this means that the gradient is a negative linear combination of the gradients of the binding (active) constraints.



Quadratic optimization with inequality constraints

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\mathbf{x}^{\mathsf{T}}\mathbf{H}\mathbf{x} + \mathbf{c}^{\mathsf{T}}\mathbf{x} + c_0 \\ \text{s.t.} & \mathbf{A}\mathbf{x} \geq \mathbf{b}. \end{array} \tag{6}$$

If $\mathbf{H}$ is positive semi-definite, then this is a convex optimization problem and we can apply Theorem 3.

Theorem 4. $\hat{\mathbf{x}}$ *is a (global) optimal solution to (6) if, and only if, there exists a vector* $\hat{\mathbf{y}} \in \mathbf{R}^m$ *such that*

(1) $\mathbf{H}\hat{\mathbf{x}} + \mathbf{c} = \mathbf{A}^{\mathsf{T}}\hat{\mathbf{y}}$

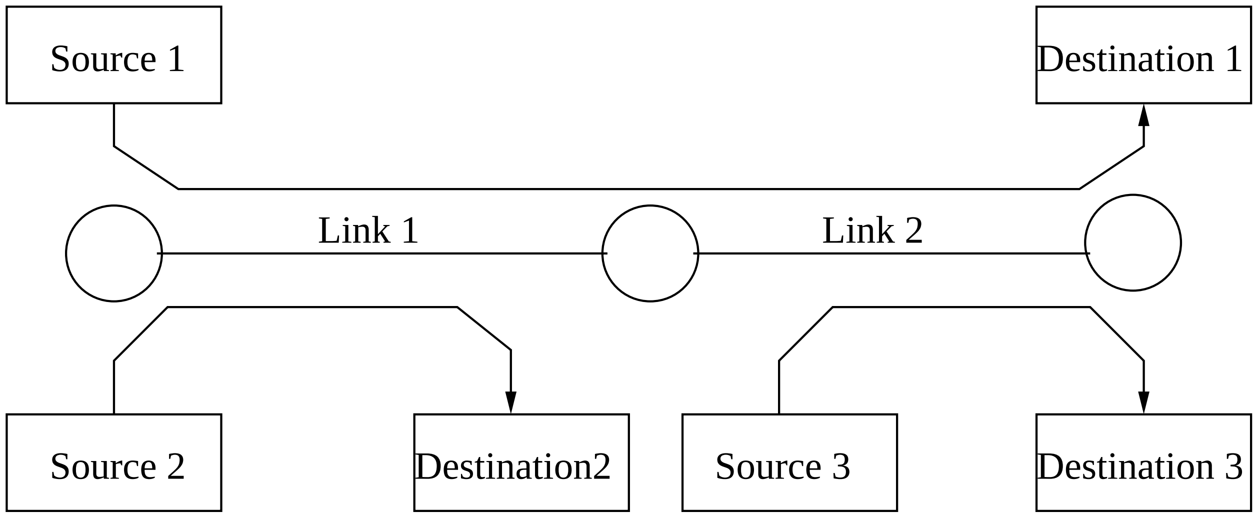
(2) $\mathbf{A}\hat{\mathbf{x}} \geq \mathbf{b}$,

(3) $\hat{\mathbf{y}} \geq 0$,

(4) $\hat{\mathbf{y}}^{\mathsf{T}}(\mathbf{A}\hat{\mathbf{x}} - \mathbf{b}) = 0$.

Traffic control in communication systems

We consider a communication network consisting of two links. Three sources are sending data over the network to three different destinations.



- Source 1 uses both links.
- Source 2 uses link 1.
- Source 3 uses link 2.

- Link 1 has capacity 2 (normalized entity [data/s])
- Link 2 has capacity 1
- The three sources sends data with speeds x_r, $r = 1, 2, 3$.
- The three sources has each a utility function $U_r(x)$, $r = 1, 2, 3$. A common choice of the utility function is $U_r(x_r) = w_r \log(x_r)$.

For efficient and fair use of the available capacity, the data speeds are chosen using the following optimization criterion:

$$\begin{aligned} \text{maximize} \quad & U_1(x_1) + U_2(x_2) + U_3(x_3) \\ \text{s.t.} \quad & x_1 + x_2 \leq 2 \\ & x_1 + x_3 \leq 1 \\ & x_1 \geq 0, \;\; x_2 \geq 0, \;\; x_3 \geq 0 \end{aligned}$$

Assume $U_k(x) = \log(x_k)$, $k = 1, 2, 3$. The optimization problem can be written

$$\begin{aligned} -\text{minimize} \quad & -\log(x_1) - \log(x_2) - \log(x_3) \\ \text{s.t.} \quad & x_1 + x_2 \leq 2 \\ & x_1 + x_3 \leq 1 \end{aligned}$$

We relaxed the constraints $x_k \geq 0$, $k = 1, \ldots, 3$ since they will be automatically satisfied, $(-\log(x) \to \infty$ as $x \to 0)$.

The optimization problem is convex since the constraints are linear inequalities and the objective function is a sum of convex functions, and hence convex.

The optimality conditions in Theorem 3 are

$$(1) \qquad \begin{aligned} -\frac{1}{x_1} + y_1 + y_2 &= 0 \\ -\frac{1}{x_2} + y_1 &= 0 \\ -\frac{1}{x_3} + y_2 &= 0 \end{aligned}$$

$$(2) \qquad \begin{aligned} x_1 + x_2 - 2 &\le 0 \\ x_1 + x_3 - 1 &\le 0 \end{aligned}$$

$$(3) \qquad \begin{aligned} y_1 &\ge 0 \\ y_2 &\ge 0 \end{aligned}$$

$$(4) \qquad \begin{aligned} y_1(x_1 + x_2 - 2) &= 0 \\ y_2(x_1 + x_3 - 1) &= 0 \end{aligned}$$

from (1) we get

$$x_1 = \frac{1}{y_1 + y_2} \qquad x_2 = \frac{1}{y_1} \qquad x_3 = \frac{1}{y_2}$$

This leads to $y_1 > 0$ and $y_2 > 0$, hence the complementarity constraint (4) shows that (2) is satisfied with equality. We get

$$\begin{aligned} \tfrac{1}{y_1+y_2} + \tfrac{1}{y_1} &= 2 \\ \tfrac{1}{y_1+y_2} + \tfrac{1}{y_2} &= 1 \end{aligned} \qquad \Rightarrow \qquad \begin{aligned} y_1 &= \tfrac{\sqrt{3}}{\sqrt{3}+1} \\ y_2 &= \sqrt{3} \end{aligned}$$

which in turn gives the optimal data speeds

$$\hat{x}_1 = \frac{\sqrt{3}+1}{3+2\sqrt{3}} \qquad \hat{x}_2 = \frac{\sqrt{3}}{\sqrt{3}+1}, \qquad \hat{x}_3 = \frac{1}{\sqrt{3}}$$

General nonlinear problems under equality constraints

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & h_i(\mathbf{x}) = 0, \;\; i = 1, \dots, m \end{aligned} \tag{7}$$

The feasible region $\mathcal{F} = \{\mathbf{x} \in \mathbf{R}^n : h_i(\mathbf{x}) = 0, \;\; i = 1, \dots, m\}$ is not convex unless the functions h_i are affine, *i.e.*, $h_i(\mathbf{x}) = \mathbf{a}_i^\mathsf{T}\mathbf{x} + b_i$. We assume that $n > m$.

We need the following technical assumption:

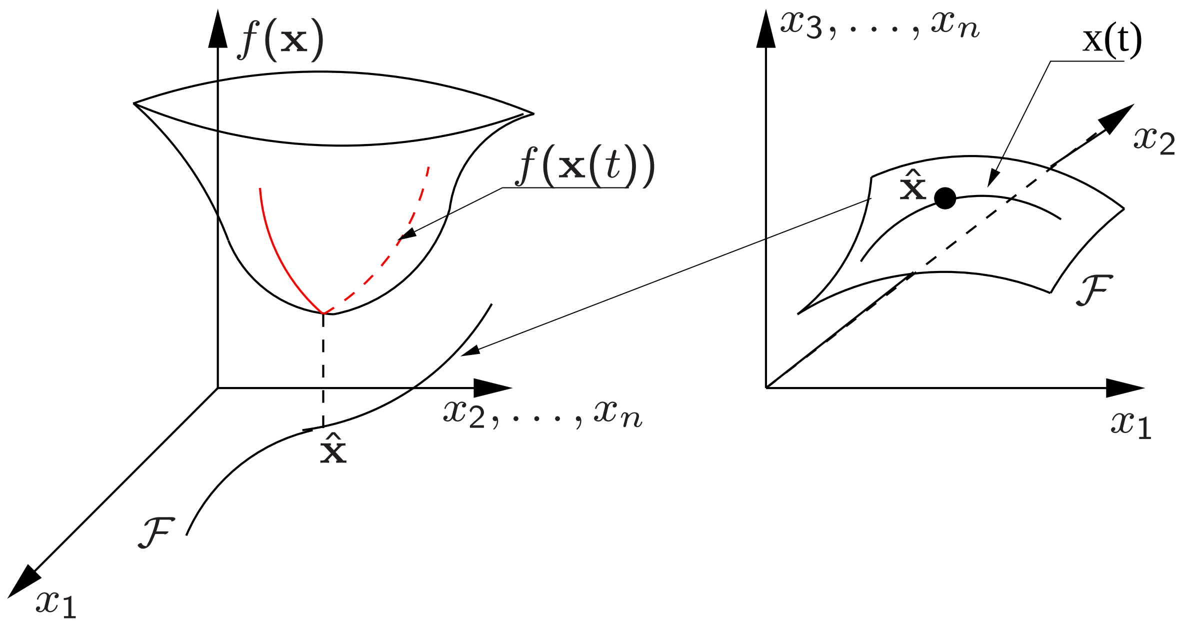
Definition 3. *A feasible solution* $\mathbf{x} \in \mathcal{F}$ *is a regular point to (7) if* $\nabla h_i(\mathbf{x})$, $i = 1, \dots, m$ *are linearly independent.*

Theorem 5 (KKT for problems with equality constraints).
Assume that $\hat{\mathbf{x}} \in \mathcal{F}$ *is a regular point and a local optimal solution to (7). 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* $\mathcal{F} = \{\mathbf{x} \in \mathbf{R}^n : g_i(\mathbf{x}) \leq 0,\ i = 1, \ldots, m\}$ *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

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

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.

General nonlinear optimization problems with inequality constraints

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & g_i(\mathbf{x}) \le 0, \;\; i = 1, \dots, m \end{aligned} \tag{8}$$

where $f(\mathbf{x})$ and $g_i(\mathbf{x})$ are real valued functions. If the problem is not convex (*i.e.*, if $\mathcal{F} = \{\mathbf{x} \in \mathbf{R}^n : g_i(\mathbf{x}) \le 0, \;\; i = 1, \dots, m\}$ is not convex and/or f is not convex on $\mathcal{F}$) it is in general only possible to derive necessary optimality conditions. The regularity condition in Definition 2 has to be replaced with a stronger condition.

Definition 4. *For $\mathbf{x} \in \mathcal{F}$ we let $\mathcal{I}_a(\mathbf{x})$ denote the index set for active constraints in the point $\mathbf{x}$, i.e., $\mathcal{I}_a(\mathbf{x}) = \{i \in \{1, \dots, m\} : g_i(\mathbf{x}) = 0\}$.*

Definition 5. *A feasible solution $\mathbf{x} \in \mathcal{F}$ is a regular point to (8) if $\nabla g_i(\mathbf{x})$ for $i \in \mathcal{I}_a(\mathbf{x})$ are linearly independent.*

Theorem 6 (KKT for general problems with inequality constraints). *Assume that $\hat{\mathbf{x}}$ is a regular point to (8) and a local optimal solution. Then there exists a vector $\hat{\mathbf{y}} \in \mathbf{R}^m$ such that*

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

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

(3) $\hat{\mathbf{y}}_i \geq 0,\ i = 1, \ldots, m,$

(4) $\hat{\mathbf{y}}_i \cdot g_i(\hat{\mathbf{x}}) = 0,\ i = 1, \ldots, m.$

Proof: *The proof is similar to the proof of Theorem 3.*