Chapter 2 Theory of Constrained Optimization

2.1 Basic notations and examples

We consider nonlinear optimization problems (NLP) of the form

$$\begin{aligned} &\text{minimize} \quad f(x) && (2.1a)\\ &\text{over} \quad x \in \mathbb{R}^n &&\\ &\text{subject to} \quad h(x) = 0 && (2.1b)\\ &\qquad\qquad\quad g(x) \le 0 \ , && (2.1c) \end{aligned}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is the objective functional and the functions $h : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^n \to \mathbb{R}^p$ describe the equality and inequality constraints.

Definition 2.1 Special cases

The minimization problem (2.1a)-(2.1c) is said to be a linear programming problem (LP) respectively a quadratic programming problem (QP), if f is linear respectively quadratic and the constraint functions h and g are affine.

Definition 2.2 Feasible set

The set of points that satisfy the equality and inequality constraints, i.e.,

$$\mathcal{F} \ := \ \{ \ x \in \mathbb{R}^n \mid h(x) \ = \ 0 \ , \ g(x) \ \le \ 0 \ \} \tag{2.2}$$

is called the feasible set of the NLP (2.1a)-(2.1c). Its elements are referred to as feasible points.

In terms of the feasible set, the NLP (2.1a)-(2.1c) can be written in the more compact form

$$\min_{x \in \mathcal{F}} f(x) \ . \tag{2.3}$$

The following examples illustrate the impact of the constraints on the solution of an NLP.

Example 2.3: Consider the constrained quadratic minimization problem

$$\begin{aligned} &\text{minimize} \quad \|x\|_2^2 && (2.4a)\\ &\text{over} \quad x \in \mathbb{R}^n &&\\ &\text{subject to} \quad g(x) := 1 - \|x\|_2^2 \le 0 \ , && (2.4b) \end{aligned}$$

where $\| \cdot \|_2$ is the Euclidean norm in $\mathbb{R}^n$.
If there is no constraint, the NLP has the unique solution $x = 0$. However, with the constraint (2.4b) any vector $x \in \mathbb{R}^n$ satisfying $\|x\|_2 = 1$ is a solution

of the NLP (2.4a)-(2.4b). Hence, if $n \geq 2$, the solution set forms an $(n-1)$-st dimensional manifold.

Example 2.4: Consider the constrained nonlinear minimization problem

$$\text{minimize} \quad (x_2 + 100)^2 + 0.01\ x_1^2 \tag{2.5a}$$
$$\text{over} \quad x = (x_1, x_2) \in \mathbb{R}^2$$
$$\text{subject to} \quad g(x) := \cos x_1 - x_2 \leq 0\ . \tag{2.5b}$$

Without constraint, the NLP has the unique solution $x = (-100, 0)^T$. With the constraint, there are infinitely many solutions near to the points

$$x \ = \ (k\pi, -1)^T \quad , \quad k = \pm 1, \pm 3, \pm 5, \ldots$$

Consequently, in contrast to the previous example, the set of solutions does not form a connected set.

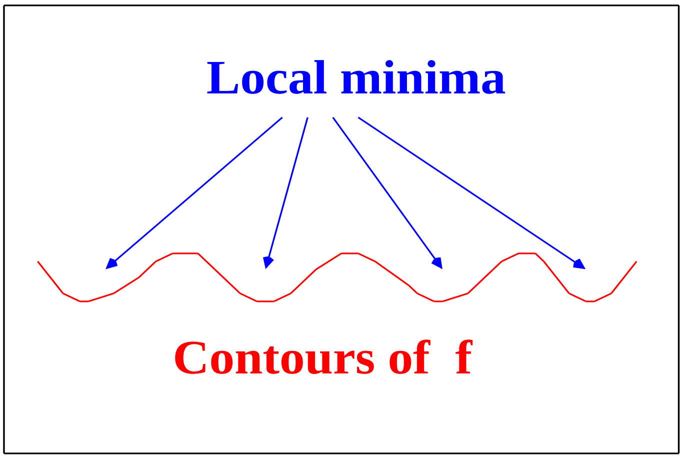

Figure 3: Constrained problem with multiple minima

The two examples give rise to the following definitions.

Definition 2.5 Local minimizers

A vector $x^* \in \mathbb{R}^n$ is called

- a local minimum of the NLP (2.1a)-(2.1c), if $x^* \in \mathcal{F}$ and there is a neighborhood $\mathcal{U}(x^*) \subset \mathbb{R}^n$ such that $f(x) \geq f(x^*)$ for all $x \in \mathcal{U}(x^*) \cap \mathcal{F}$,
- an isolated local minimum of the NLP (2.1a)-(2.1c), if $x^* \in \mathcal{F}$ and there is a neighborhood $\mathcal{U}(x^*) \subset \mathbb{R}^n$ such that x^* is the only local minimum in $\mathcal{U}(x^*) \cap \mathcal{F}$.

We will now focus our interest on the characterization of solutions to the NLP (2.1a)-(2.1c).

Example 2.6 A single equality constraint
Consider the NLP

$$\text{minimize} \quad x_1 + x_2 \tag{2.6a}$$
$$\text{over} \quad x = (x_1, x_2) \in \mathbb{R}^2$$
$$\text{subject to} \quad h(x) := x_1^2 + x_2^2 - 2 = 0 \ . \tag{2.6b}$$

The unique solution of (2.6a)-(2.6b) is obviously given by $x^* = (-1,-1)^T$. Computing the gradients of f and h in x^*, we obtain

$$\nabla f(x^*) \ = \ (1,1)^T \quad , \quad \nabla h(x^*) \ = \ (-2,-2)^T \ .$$

Obviously, $\nabla f(x^*)$ and $\nabla h(x^*)$ are parallel, i.e., there exists a scalar $\lambda^* \in \mathbb{R}$, in this particular case $\lambda^* = 1/2$ such that

$$\nabla f(x^*) \ = \ - \ \lambda^* \ \nabla h(x^*) \ . \tag{2.7}$$

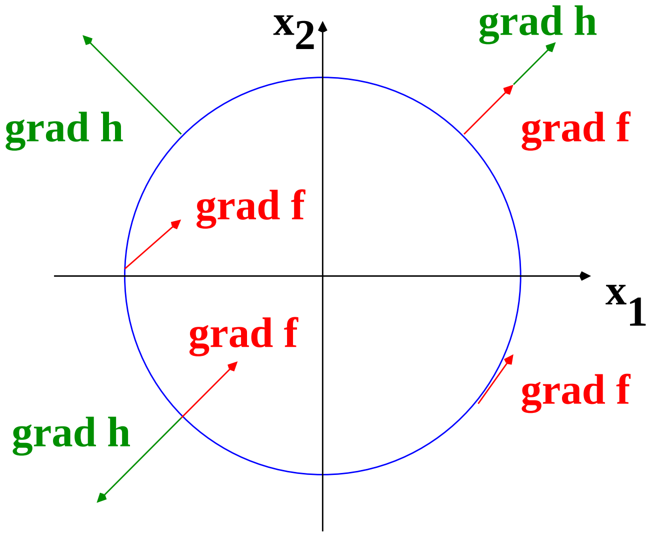

Figure 4: Function and constraint gradients in Example 2.6

We will now show that (2.7) is a necessary condition for optimality in the general case.
Assume that $x \in \mathcal{F}$. Then, Taylor expansion of $h(x+d)$, $d \in \mathbb{R}^n$, gives

$$h(x+d) \ \approx \ \underbrace{h(x)}_{=\,0} \ + \nabla h(x)^T \ d \ .$$

If we want to retain feasibility at $x + d$, we have to require

$$\nabla h(x)^T \, d \; = \; 0 \; . \tag{2.8}$$

On the other hand, if we want that the direction d results in a decrease of the objective functional f, there must hold

$$0 \; > \; f(x+d) \; - \; f(x) \; \approx \; \nabla f(x)^T \, d \; ,$$

which leads to the requirement

$$\nabla f(x)^T \, d \; < \; 0 \; . \tag{2.9}$$

Consequently, if x is a local minimum, there is no direction d satisfying (2.8) and (2.9) simultaneously. The only way that such a direction does not exist is that $\nabla f(x)$ and $\nabla h(x)$ are parallel.

Introducing the Lagrangian functional $\mathcal{L} : \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}$ by means of

$$\mathcal{L}(x, \lambda) \; := \; f(x) \; + \; \lambda \, h(x) \; , \tag{2.10}$$

and observing $\nabla_x \mathcal{L}(x, \lambda) = \nabla f(x) + \lambda \nabla h(x)$, condition (2.7) can be equivalently stated as

$$\nabla_x \mathcal{L}(x^*, \lambda^*) \; = \; 0 \; . \tag{2.11}$$

Example 2.7 A single inequality constraint

Consider the inequality constrained NLP

$$\text{minimize} \quad x_1 + x_2 \tag{2.12a}$$

$$\text{over} \quad x = (x_1, x_2) \in \mathbb{R}^2$$

$$\text{subject to} \quad g(x) := x_1^2 + x_2^2 - 2 \le 0 \; . \tag{2.12b}$$

The feasible region is the closed ball $\bar{B}_{\sqrt{2}}(0)$ with radius $\sqrt{2}$ around the origin. As in the previous example, the solution is $x^* = (-1, -1)^T$.

Again, a feasible point $x \in \mathcal{F}$ is not optimal, if we can find a direction $d \in \mathbb{R}^2$ such that d is a descent direction, i.e.,

$$\nabla f(x)^T d \; < \; 0 \; , \tag{2.13}$$

and $x + d$ is still feasible, i.e.,

$$0 \; \ge \; g(x+d) \; \approx \; g(x) \; + \; \nabla g(x)^T d \; .$$

To first order, this leads to the condition

$$g(x) \; + \; \nabla g(x)^T d \; \le \; 0 \; . \tag{2.14}$$

For the characterization of directions d that satisfy (2.13) and (2.14), we distinguish the two cases $x \in B_{\sqrt{2}}(0)$ and $x \in \partial B_{\sqrt{2}}(0)$.

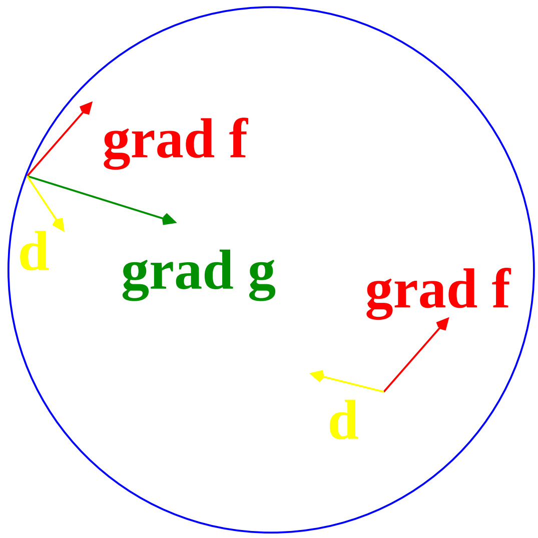

Figure 5: Descent directions from two feasible points in Example 2.7 where the constraint is active/inactive

Case I: $x \in B_{\sqrt{2}}(0)$
In this case we have $g(x) < 0$. Obviously, (2.14) is satisfied for any vector d provided its length is sufficiently small. In particular, for $\nabla f(x) \neq 0$, a direction $d \neq 0$ that satisfies (2.13) and (2.14) is given by

$$d \;=\; \alpha\, g(x)\, \frac{\nabla f(x)}{\|\nabla f(x)\|_2} \quad , \quad 0 < \alpha < \begin{cases} +\infty & , \quad \nabla g(x) = 0 \\ \frac{1}{\|\nabla g(x)\|_2} & , \quad \nabla g(x) \neq 0 \end{cases} .$$

The only situation where such a direction fails to exist is when

$$\nabla f(x) \;=\; 0 \; . \tag{2.15}$$

Case II: $x \in \partial B_{\sqrt{2}}(0)$
In this case we have $g(x) = 0$ and hence, conditions (2.13) and (2.14) reduce to

$$\nabla f(x)^T d \;<\; 0 \quad , \quad \nabla g(x)^T d \;\leq\; 0 \; .$$

It is obvious that these conditions cannot be satisfied if $\nabla f(x)$ and $\nabla g(x)$ point in different directions, i.e.,

$$\nabla f(x) \;=\; -\,\mu\, \nabla g(x) \quad \text{for some } \mu \geq 0 \; . \tag{2.16}$$

Note that the sign of the Lagrangian multiplier μ is essential.
Again, the optimality condition for both cases can be summarized by considering the Lagrangian functional

$$\mathcal{L}(x,\mu) \ := \ f(x) \ + \ \mu \ g(x) \ .$$

When no first-order feasible descent direction exists at some $x^* \in \mathcal{F}$, we have

$$\nabla_x \mathcal{L}(x^*,\mu^*) \ = \ 0 \quad \text{for some } \mu^* \geq 0 \ , \tag{2.17}$$

and we also require

$$\mu^* \ g(x^*) \ = \ 0 \ . \tag{2.18}$$

Condition (2.18) is called a complementarity condition. It says that the Lagrangian multiplier μ^* can be strictly positive only if the constraint g is active, i.e., $g(x^*) = 0$.
In case I, we have $g(x^*) < 0$ so that according to (2.18) the multiplier has to satisfy $\mu^* = 0$. Hence, (2.17) reduces to $\nabla f(x^*) = 0$ which corresponds to (2.15).
On the other hand, in case II the multiplier μ^* can take a nonnegative value. Consequently, (2.17) is equivalent to (2.16).

2.2 First Order Optimality Conditions

The Lagrangian associated with the NLP (2.1a)-(2.1c) is given by the functional $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p_+ \to \mathbb{R}$

$$\mathcal{L}(x,\lambda,\mu) \ := \ f(x) \ + \ \lambda^T h(x) \ + \ \mu^T g(x) \ . \tag{2.19}$$

Definition 2.8 Active set
Let $x \in \mathcal{F}$. Then, the set

$$\mathcal{I}_{ac}(x) \ := \ \{ \ 1 \leq i \leq p \mid g_i(x) = 0 \ \} \tag{2.20}$$

is said to be the set of active inequality constraints at x. Its complement $\mathcal{I}_{ia}(x) := \{1,...,p\} \setminus \mathcal{I}_{ac}(x)$ is referred to as the set of inactive inequality constraints.

The vectors $\nabla h_i(x)$, $1 \leq i \leq m$, and $\nabla g_i(x)$, $1 \leq i \leq p$, are called the normals of the equality constraints h_i respectively g_i at x.
Considering Example 2.6, if we choose

$$h(x) \ = \ (x_1^2 \ + \ x_2^2 \ - \ 2)^2 \ ,$$

we obtain

$$\nabla h(x) \ = \ 4 \ (x_1^2 \ + \ x_2^2 \ - \ 2) \ (x_1,x_2)^T \ .$$

Obviously, $\nabla h(x) = 0$ for all $x \in \mathcal{F} = \partial B_{\sqrt{2}}(0)$. Consequently, $\nabla f(x) = \lambda \nabla h(x)$ does no longer hold true at the minimum $x^* = (-1,-1)^T$. In the sequel, we want to exclude such a degenerate behavior by the following constraint qualification.

Definition 2.9 Linear Independence Constraint Qualification

Let $\mathcal{I}_{ac}(x^*)$, $x^* \in \mathcal{F}$, be the set of active inequality constraints. Then, the Linear Independence Constraint Qualification (LICQ) is satisfied at x^*, if the set of active constraint gradients

$$\{ \ \nabla h_1(x^*), ..., \nabla h_m(x^*), \nabla g_i(x^*), i \in \mathcal{I}_{ac}(x^*) \ \} \tag{2.21}$$

is linearly independent.

We further introduce the concepts of feasible sequences and associated limiting directions.

Definition 2.10 Feasible sequence and limiting direction

(i) A sequence $\{x_k\}_{\mathbb{N}}$ with $x_k \in \mathbb{R}^n, k \in \mathbb{N}$, is said to be a feasible sequence with respect to a feasible point $x^* \in \mathbb{R}^n$, if there exists $k_0 \in \mathbb{N}$ such that the following properties are satisfied

$$x_k \ \neq \ x^* \ , \ k \in \mathbb{N} \ , \tag{2.22a}$$

$$\lim_{k\to\infty} x_k \ = \ x^* \ , \tag{2.22b}$$

$$x_k \in \mathcal{F} \ , \ k \geq k_0 \ . \tag{2.22c}$$

The set of all feasible sequences with respect to a feasible point $x^* \in \mathcal{F}$ will be denoted by $\mathcal{T}(x^*)$.

(ii) A vector $d \in \mathbb{R}^n$ is said to be a limiting direction of a feasible sequence $\{x_k\}_{\mathbb{N}} \in \mathcal{T}(x^*)$, if there exists a subsequence $\{x_k\}_{\mathbb{N}'}, \mathbb{N}' \subset \mathbb{N}$, such that

$$\lim_{k\in\mathbb{N}'} \frac{x_k - x^*}{\|x_k - x^*\|_2} \ = \ d \ . \tag{2.23}$$

Definition 2.11 Local solution

A feasible point $x^* \in \mathcal{F}$ is called a local solution of the NLP (2.1a)-(2.1c), if for all feasible sequences $\{x_k\}_{\mathbb{N}} \in \mathcal{T}(x^*)$ there exists $k_0 \in \mathbb{N}$ such that

$$f(x_k) \ \geq \ f(x^*) \ , \ k \geq k_0 \ .$$

Example 2.12: Let us consider the equality constrained NLP (2.6a)-(2.6b) and the feasible point $x^* = (-\sqrt{2}, 0)^T$.

A feasible sequence is given by

$$x_k \ = \ \begin{pmatrix} -\sqrt{2 - 1/k^2} \\ 1/k \end{pmatrix} \ , \ k \in \mathbb{N} \ .$$

Obviously, $f(x_k) \geq f(x^*)$, $k \in \mathbb{N}$, and $d = (0,1)^T$ is the limiting direction of that feasible sequence.
On the other hand, another feasible sequence is

$$x_k \;=\; \begin{pmatrix} -\sqrt{2 - 1/k^2} \\ -1/k \end{pmatrix} \;,\; k \in \mathbb{N} \;.$$

In this case we have $f(x_k) < f(x^*)$, $k \geq 2$, and $d = (0,-1)^T$ is the limiting direction.
It follows that x^* cannot be a local solution of (2.6a)-(2.6b).

Theorem 2.12 Characterization of local solutions
If $x^* \in \mathcal{F}$ is a local solution of the NLP (2.1a)-(2.1c), then for any feasible sequence $\{x_k\}_{\mathbb{N}} \in \mathcal{T}(x^*)$ and any limiting direction d of the feasible sequence there holds

$$\nabla f(x^*)^T d \;\geq\; 0 \;. \tag{2.24}$$

Proof: The proof is by contradiction. We assume that there exists a feasible sequence $\{x_k\}_{\mathbb{N}} \subset \mathcal{T}(x^*)$ and a limiting direction d such that

$$\nabla f(x^*)^T d \;<\; 0 \;. \tag{2.25}$$

Let $\mathbb{N}' \subset \mathbb{N}$ be such that $\lim\limits_{k \in \mathbb{N}'} x_k = x^*$. By Taylor expansion, for $k \in \mathbb{N}'$ we obtain

$$\begin{aligned} f(x_k) \;&=\; f(x^*) \;+\; \nabla f(x^*)^T (x_k - x^*) \;+\; o(\|x_k - x^*\|_2) \;= \\ &=\; f(x^*) \;+\; \|x_k - x^*\|_2 \; \nabla f(x^*)^T d \;+\; o(\|x_k - x^*\|_2) \;. \end{aligned}$$

In view of (2.25) there exists $k_0 \in \mathbb{N}'$ such that

$$f(x_k) \;<\; f(x^*) \;+\; \frac{1}{2} \; \|x_k - x^*\|_2 \; \nabla f(x^*)^T d \;,\; k \geq k_0 \;,$$

whence $f(x_k) < f(x^*)$, $k \geq k_0$. Consequently, x^* is not a local solution.

•

The Linear Independence Constraint Qualification (LICQ) allows the characterization of the set of all possible limiting directions d of a feasible sequence in $\mathcal{T}(x^*)$ in terms of the gradients $\nabla h_i(x^*)$, $1 \leq i \leq m$, and $\nabla g_i(x^*)$, $i \in \mathcal{I}_{ac}(x^*)$.

Lemma 2.13 Characterization of limiting directions
(i) If $d \in \mathbb{R}^n$ is a limiting direction of a feasible sequence in $\mathcal{T}(x^*)$, then

$$\nabla h_i(x^*)^T d \;=\; 0 \;,\; 1 \leq i \leq m \;, \tag{2.26a}$$

$$\nabla g_i(x^*)^T d \;\leq\; 0 \;,\; i \in \mathcal{I}_{ac}(x^*) \;. \tag{2.26b}$$

(ii) On the other hand, if (2.26a)-(2.26b) holds true with $\|d\|_2 = 1$ and if the LICQ is satisfied, then d is a limiting direction of the feasible sequence.

Proof: For the proof of part (i) let $\{x_k\}_{\mathbb{N}} \in \mathcal{T}(x^*)$ a feasible sequence with limiting direction d such that for $\mathbb{N}' \subset \mathbb{N}$

$$\lim_{k \in \mathbb{N}'} \frac{x_k - x^*}{\|x_k - x^*\|_2} \; = \; d \; .$$

It follows that

$$x_k \; = \; x^* \; + \; d \; \|x_k - x^*\|_2 \; + \; o(\|x_k - x^*\|_2) \; .$$

For $1 \leq i \leq m$, by Taylor expansion we obtain

$$\begin{aligned}
0 \; &= \; \frac{1}{\|x_k - x^*\|_2} \; h_i(x_k) \; = \\
&= \; \frac{1}{\|x_k - x^*\|_2} \; \big[h_i(x^*) \; + \; \|x_k - x^*\|_2 \; \nabla h_i(x^*)^T d \; + \; o(\|x_k - x^*\|_2)\big] \; = \\
&= \; \nabla h_i(x^*)^T d \; + \; \frac{o(\|x_k - x^*\|_2)}{\|x_k - x^*\|_2} \; .
\end{aligned}$$

For $k \to \infty$ this gives $\nabla h_i(x^*)^T d = 0$.
Similarly, for $i \in \mathcal{I}_{ac}(x^*)$ we get

$$\begin{aligned}
0 \; &\geq \; \frac{1}{\|x_k - x^*\|_2} \; g_i(x_k) \; = \\
&= \; \frac{1}{\|x_k - x^*\|_2} \; \big[g_i(x^*) \; + \; \|x_k - x^*\|_2 \; \nabla g_i(x^*)^T d \; + \; o(\|x_k - x^*\|_2)\big] \; = \\
&= \; \nabla g_i(x^*)^T d \; + \; \frac{o(\|x_k - x^*\|_2)}{\|x_k - x^*\|_2} \; .
\end{aligned}$$

Hence, $k \to \infty$ implies $\nabla g_i(x^*)^T d \leq 0$.
For the proof of (ii) let $p^* \; := \;$ card $\mathcal{I}_{ac}(x^*)$ and reorder $g_1, ..., g_p$ such that $g_i(x^*) = 0 \; , \; 1 \leq i \leq p^*$. Set $q^* := m + p^*$ and consider the matrix $A \in \mathbb{R}^{q^* \times n}$

$$A \; = \; \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} \quad , \quad \begin{array}{l} A_1 \; = \; \begin{bmatrix} \nabla h_1(x^*) \\ \cdot \\ \nabla h_m(x^*) \end{bmatrix} \\ \\ A_2 \; = \; \begin{bmatrix} \nabla g_1(x^*) \\ \cdot \\ \nabla g_{p^*}(x^*) \end{bmatrix} \end{array} \; . \tag{2.27}$$

If LIQC holds true, the matrix A has full row rank q^*. We denote by Z the matrix whose columns form a basis of the kernel of A, i.e.,

$$Z \in \mathbb{R}^{n \times (n - q^*)} \text{ has full column rank} \quad , \quad AZ \; = \; 0 \; . \tag{2.28}$$

Let d be a limiting direction of a feasible sequence satisfying (2.26a),(2.26b) and let $\{t_k\}_{\mathbb{N}}$ be a null sequence of positive real numbers. Consider the parametrized system of equations

$$R(z,t) := \begin{bmatrix} h_i(z) - tA_1 d \\ g_i(z) - tA_2 d \\ Z^T(z - x^* - td \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad \begin{matrix} 1 \le i \le m \\ 1 \le i \le p^* \end{matrix} \quad . \tag{2.29}$$

We show that there exists $k_0 \in \mathbb{N}$ such that for $t = t_k$, $k \ge k_0$, the system (2.29) admits a unique solution $z = x_k$ and that $\{x_k\}_{\mathbb{N}}$ is a feasible sequence with limiting direction d:
For $t = 0$, the solution of (2.29) is given by $z = x^*$ with the Jacobian

$$\nabla_z R(x^*, 0) = \begin{bmatrix} A \\ Z^T \end{bmatrix} .$$

By construction of Z it follows that $\nabla_z R(x^*, 0) \in \mathbb{R}^{n\times n}$ is regular. The implicit function theorem implies that (2.29) is uniquely solvable for sufficiently small t_k, i.e., there exists $k_0 \in \mathbb{N}$ such that for $k \ge k_0$ and $t = t_k$ the system (2.29) has a unique solution $z = x_k$. Observing (2.26a),(2.26b), and (2.29), we get

$$h_i(x_k) = t_k \nabla h_i(x_k)^T d = 0 \ , \ 1 \le i \le m \ , \tag{2.30a}$$

$$g_i(x_k) = t_k \nabla g_i(x_k)^T d \le 0 \ , \ 1 \le i \le p^* \ , \tag{2.30b}$$

which proves that x_k is feasible.
We now show that $x_k = z(t_k) \neq x^*$ for all k. The proof is by contradiction: Assume $z(\bar t) = x^*$ for some $\bar t > 0$, i.e.,

$$\begin{bmatrix} h_i(x^*) - \bar t A_1 d \\ g_i(x^*) - \bar t A_2 d \\ - Z^T(\bar t d) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad \begin{matrix} 1 \le i \le m \\ 1 \le i \le p^* \end{matrix} \quad . \tag{2.31}$$

Since $h_i(x^*) = 0$, $1 \le i \le m$, $g_i(x^*) = 0$, $1 \le i \le p^*$, and $\begin{bmatrix} A \\ Z^T \end{bmatrix}$ is regular, (2.31) has the unique solution $d = 0$ contradicting $\|d\|_2 = 1$.
It remains to be shown that d is a limiting direction of the feasible sequence $\{x_k\}_{\mathbb{N}}$. Observing $R(x_k, t_k) = 0$, by Taylor expansion we find

$$\begin{aligned} 0 = R(x_k, t_k) &= \begin{bmatrix} h_i(x_k) - t_k A_1 d \\ g_i(x_k) - t_k A_2 d \\ Z^T(x_k - x^* - t_k d) \end{bmatrix} = \\ &= \begin{bmatrix} A_1(x_k - x^*) + o(\|x_k - x^*\|_2) - t_k A_1 d \\ A_2(x_k - x^*) + o(\|x_k - x^*\|_2) - t_k A_2 d \\ Z^T(x_k - x^* - t_k d) \end{bmatrix} = \\ &= \begin{bmatrix} A \\ Z^T \end{bmatrix} (x_k - x^* - t_k d) + o(\|x_k - x^*\|_2) . \end{aligned}$$

Setting

$$d_k \ := \ \frac{x_k - x^*}{\|x_k - x^*\|_2}$$

and observing that $\begin{bmatrix} A \\ Z^T \end{bmatrix}$ is nonsingular, it follows that

$$\lim_{k\to\infty} \left[\, d_k \ - \ \frac{t_k}{\|x_k - x^*\|_2} \, d \, \right] \ = \ 0 \ .$$

Since $\|d_k\|_2 = 1 \ , \ k \in \mathbb{N}$, and $\|d\|_2 = 1$, we deduce

$$\lim_{k\to\infty} \frac{t_k}{\|x_k - x^*\|_2} \ = \ 1 \ ,$$

whence $\lim\limits_{k\to\infty} d_k = d$.

•

Definition 2.14 Tangent and normal cone to the feasible set
A vector $w \in \mathbb{R}^n$ is tangent to the feasible set $\mathcal{F}$ at $x \in \mathcal{F}$, if for all sequences $\{x_k\}_{\mathbb{N}} \ , \ x_k \in \mathcal{F} \ , \ k \in \mathbb{N}$, with $\lim\limits_{k\to\infty} x_k = x$ and all null sequences $\{t_k\}_{\mathbb{N}}$ of positive real numbers $t_k \ , \ k \in \mathbb{N}$, there exists a sequence $\{w_k\}_{\mathbb{N}} \ , \ w_k \in \mathbb{R}^n \ , \ k \in \mathbb{N}$, with $\lim\limits_{k\to\infty} w_k = w$ such that $x_k + t_k w_k \in \mathcal{F} \ , \ k \in \mathbb{N}$.
The set $T_{\mathcal{F}}(x)$ of all tangent vectors to $\mathcal{F}$ at $x \in \mathcal{F}$ is a cone, i.e., it has the property

$$w \in T_{\mathcal{F}}(x) \quad \Longrightarrow \quad \alpha w \in T_{\mathcal{F}}(x) \text{ for all } \alpha \geq 0 \ .$$

$T_{\mathcal{F}}(x)$ is called the tangent cone to $\mathcal{F}$ at x.
The orthogonal complement to the tangent cone

$$N_{\mathcal{F}}(x) \ := \ \{ \ v \in \mathbb{R}^n \mid v^T w \ \leq \ 0 \ , \ w \in T_{\mathcal{F}}(x) \ \}$$

is called the normal cone to the feasible set $\mathcal{F}$ at $x \in \mathcal{F}$.

Lemma 2.15 Limiting directions and the tangent cone
Given $x^* \in \mathcal{F}$, the set

$$F_1 \ := \ \left\{ \ \alpha d \mid \alpha \geq 0 \ , \ \begin{array}{lcl} \nabla h_i(x^*)^T d \ = \ 0 & , & 1 \leq i \leq m \ , \\ \nabla g_i(x^*)^T d \ \leq \ 0 & , & i \in \mathcal{I}_{ac}(x^*) \end{array} \ \right\} \tag{2.32}$$

is a cone. If LICQ is satisfied, F_1 is the tangent cone to the feasible set $\mathcal{F}$ at x^*.
Proof: The proof is left as an exercise.

•

We now show that the nonexistence of a descent direction for the objective functional f can be stated in terms of the normal cone to the feasible set which is generated by the gradients of the active constraints.

Lemma 2.16 Nonexistence of a descent direction
There is no descent direction $d \in F_1$, i.e., satisfying $\nabla f(x^*)^T d < 0$, if and only if there exist a vector $\lambda \in \mathbb{R}^m$ and a vector $\mu \in \mathbb{R}_+^{p^*}$ such that

$$\nabla f(x^*) \in N := \{s \mid s = \sum_{i=1}^m \lambda_i \nabla h_i(x^*) - \sum_{i \in \mathcal{I}_{ac}(x^*)} \mu_i \nabla g_i(x^*)\} \ . \tag{2.33}$$

Proof: We first prove the sufficiency of (2.33). For that purpose, we assume that (2.33) holds true and $d \in F_1$. We obtain

$$\nabla f(x^*)^T d \ = \ \sum_{i=1}^m \lambda_i \underbrace{\nabla h_i(x^*)^T d}_{=\ 0} \ - \sum_{i \in \mathcal{I}_{ac}(x^*)} \mu_i \underbrace{\nabla g_i(x^*)^T d}_{\leq\ 0} \ \geq \ 0 \ .$$

The proof of the necessity of (2.33) is by contradiction. If $\nabla f(x^*) \notin N$, then we can find a vector $d \in F_1$ for which $\nabla f(x^*)^T d < 0$.
Let $\hat{s} \in N$ be the best approximation of $\nabla f(x^*)$ in N, i.e.,

$$\|\hat{s} \ - \ \nabla f(x^*)\|_2 \ = \ \min_{s \in N} \|s \ - \ \nabla f(x^*)\|_2 \ . \tag{2.34}$$

Since $t\hat{s} \in N$ for all $t \geq 0$ and $\|t\hat{s} - \nabla f(x^*)\|_2$ is minimized at $t = 1$, there holds

$$\frac{d}{dt} \ \|t\hat{s} - \nabla f(x^*)\|_2|_{t=1} \ = \ 0 \tag{2.35}$$
$$\Longrightarrow \ (-2\hat{s}^T \nabla f(x^*) + 2t\hat{s}^T\hat{s})|_{t=1} \ = \ 0 \quad \Longrightarrow \quad \hat{s}^T(\hat{s} - \nabla f(x^*)) \ = \ 0 \ .$$

Now, let $s \in N$, $s \neq \hat{s}$. Due to the convexity of N and the minimizing property of $\hat{s}$ we have

$$\|\hat{s} \ + \ \theta(s - \hat{s}) \ - \ \nabla f(x^*)\|_2^2 \ \geq \ \|\hat{s} \ - \ \nabla f(x^*)\|_2^2 \quad , \quad \theta \in [0,1] \ ,$$

from which we readily deduce

$$2\theta(s - \hat{s})^T(\hat{s} - \nabla f(x^*)) \ + \ \theta^2 \|s - \hat{s}\|_2^2 \ \geq \ 0 \ .$$

Division by $\theta \neq 0$ and the limit process $\theta \to 0$ yield $(s - \hat{s})^T(\hat{s} - \nabla f(x^*)) \geq 0$. Observing (2.34), we get

$$s^T(\hat{s} - \nabla f(x^*)) \ \geq \ 0 \quad , \quad s \in N \ . \tag{2.36}$$

Setting $d := \hat{s} - \nabla f(x^*)$, we will show that $d \in F_1$ is a descent direction, i.e., $\nabla f(x^*)^T d < 0$. We first note that $d \neq 0$, since $\nabla f(x^*) \notin N$. It follows from (2.35) that

$$\nabla f(x^*)^T d \ = \ (\hat{s} - d)^T d \ = \ \hat{s}^T(\hat{s} - \nabla f(x^*)) \ - \ d^T d \ = \ - \ \|d\|_2^2 \ < \ 0 \ .$$

It remains to be shown that $d \in F_1$, i.e., that (2.26a) and (2.26b) are satisfied: Choosing $\lambda_i = \pm\delta_{ij}$, $1 \le i \le m$, and $\mu_i = \delta_{ij}$, $i \in \mathcal{I}_{ac}(x^*)$, in the definition of the cone N (cf. (2.33)), we find

$$\begin{aligned} \pm\nabla h_i(x^*) &\in N \quad , \quad 1 \le i \le m \ , \\ -\nabla g_i(x^*) &\in N \quad , \quad i \in \mathcal{I}_{ac}(x^*) \ . \end{aligned}$$

Hence, substituting $\hat{s} - \nabla f(x^*)$ by d in (2.36) and choosing $s = \pm\nabla h_i(x^*)$, $1 \le i \le m$, respectively $s = -\nabla g_i(x^*)$, $i \in \mathcal{I}_{ac}(x^*)$, we arrive at

$$\pm\nabla h_i(x^*)^T d \ \ge \ 0 \quad \Longrightarrow \quad \begin{aligned} \nabla h_i(x^*)^T d &= 0 \quad , \quad 1 \le i \le m \ , \\ \nabla g_i(x^*)^T d &\le 0 \quad , \quad i \in \mathcal{I}_{ac}(x^*) \ . \end{aligned}$$

•

We are now in a position to state the main result of this section, the first order necessary optimality conditions also known as the Karush-Kuhn-Tucker (KKT) conditions.

Theorem 2.17 Karush-Kuhn-Tucker conditions

Assume that $x^* \in \mathcal{F}$ is a local solution of (2.1a)-(2.1c) and that the LICQ is satisfied at x^*. Then, there exist Lagrange multipliers $\lambda^* \in \mathbb{R}^m$ and $\mu^* \in \mathbb{R}^p$ such that the following conditions hold true at (x^*, λ^*, μ^*):

$$\nabla_x \mathcal{L}(x^*, \lambda^*, \mu^*) \ = \ 0 \ , \tag{2.37a}$$

$$h_i(x^*) \ = \ 0 \ , \ 1 \le i \le m \ , \tag{2.37b}$$

$$g_i(x^*) \ \le \ 0 \ , \ 1 \le i \le p \ , \tag{2.37c}$$

$$\mu_i^* \ \ge \ 0 \ , \ 1 \le i \le p \ , \tag{2.37d}$$

$$\mu_i^* g_i(x^*) \ = \ 0 \ , \ 1 \le i \le p \ . \tag{2.37e}$$

Proof: We first prove that there exist multipliers μ_i , $i \in \mathcal{I}_{ac}(x^*)$, such that (2.33) in Lemma 2.16 is satisfied. From Theorem 2.12 we know that $\nabla f(x^*)^T d \ge 0$ for all limiting directions d of feasible sequences. Under the condition LICQ, Lemma 2.13 states that the set of all possible limiting directions satisfies (2.26). Consequently, all directions d for which (2.26) holds true, also satisfy $\nabla f(x^*)^T d \ge 0$. Hence, by Lemma 2.16 there exists $\mu \in \mathbb{R}^p_+$ such that (2.33) is fulfilled. We now define $\lambda^* \in \mathbb{R}^m$ and $\mu^* \in \mathbb{R}^p$ according to

$$\lambda_i^* \ := \ 0 \ , \ 1 \le i \le m \quad , \quad \mu_i^* \ := \ \begin{cases} \mu_i & , \quad i \in \mathcal{I}_{ac}(x^*) \\ 0 & , \quad i \in \mathcal{I}_{in}(x^*) \end{cases} \ . \tag{2.38}$$

With that definition of the multipliers λ^*, μ^*, taking advantage of (2.33) we have

$$\mathcal{L}_x(x^*, \lambda^*, \mu^*) \ = $$

$$\nabla f(x^*) + \underbrace{\sum_{i=1}^m \lambda_i^* \nabla h_i(x^*)}_{=\,0} + \sum_{i \in \mathcal{I}_{ac}(x^*)} \mu_i^* \nabla g_i(x^*) + \underbrace{\sum_{i \in \mathcal{I}_{in}(x^*)} \mu_i^* \nabla g_i(x^*)}_{=\,0} \ = \ 0 \ ,$$

which is (2.37a) of the KKT-conditions.
Since x^* is feasible, it is obvious that (2.37b) and (2.37c) are satisfied.
In view of (2.33) we have $\mu_i^* \geq 0 \ , \ i \in \mathcal{I}_{ac}(x^*)$, whereas $\mu_i^* = 0 \ , \ i \in \mathcal{I}_{in}(x^*)$ by (2.38). Hence, $\mu_i^* \geq 0 \ , \ 1 \leq i \leq p$, which is (2.37d).
Finally, $g_i(x^*) = 0 \ , \ i \in \mathcal{I}_{ac}(x^*)$, and $\mu_i^* = 0 \ , \ i \in \mathcal{I}_{in}(x^*)$ so that $\mu_i^* g_i(x^*) = 0 \ , \ 1 \leq i \leq p$, which is (2.37e).

•

Definition 2.18 Strict complementarity
Let $x^* \in \mathcal{F}$ be a local solution of the NLP (2.1a)-(2.1c) and $\lambda^* \ , \ \mu^*$ Lagrange multipliers satisfying the KKT conditions (2.37a)-(2.37e). Then, strict complementarity holds true if

$$\mu_i^* \ > \ 0 \quad , \quad i \in \mathcal{I}_{ac}(x^*) \ . \tag{2.39}$$

Remark 2.19 Since $\mu_i^* = 0 \ , \ i \in \mathcal{I}_{in}(x^*)$, strict complementarity means that exactly one of the quantities μ_i^* and $g_i(x^*)$ is zero for each $1 \leq i \leq p$.

2.3 Second order optimality conditions

The first order optimality conditions (KKT-conditions) give information how the first derivatives of the objective functional f and the constraints h and g behave at a local solution $x^* \in \mathcal{F}$ of the NLP (2.1a)-(2.1c). If we proceed from x^* along a vector $w \in F_1$, then the first order approximation $f(x^*) + \nabla f(x^*)^T w$ of $f(x^* + w)$ either increases ($\nabla f(x^*)^T w > 0$) or stays constant ($\nabla f(x^*)^T w = 0$). In the latter case, additional information will be provided by the second derivatives of f, h and g at x^*. In the sequel, we assume f, h and g to be twice continuously differentiable.
Given Lagrange multipliers $\lambda^* \in \mathbb{R}^m \ , \mu^* \in \mathbb{R}^p$ that satisfy the KKT-conditions (2.37a)-(2.37e), we define a subset

$$F_2(\lambda^*, \mu^*) \subset F_1$$

as follows

$$F_2(\lambda^*, \mu^*) := \{ w \in F_1 \mid \nabla g_i(x^*)^T w = 0 \ , \ i \in \mathcal{I}_{ac}(x^*) \text{ with } \mu_i^* > 0 \} \ . \tag{2.40}$$

By the definition of F_1 (cf. (2.32)) we have

$$w \in F_2(\lambda^*, \mu^*) \quad \Longleftrightarrow \tag{2.41}$$

$$\left\{ \begin{array}{lcll} \nabla h_i(x^*)^T w & = & 0 \quad , & 1 \leq i \leq m \ , \\ \nabla g_i(x^*)^T w & = & 0 \quad , & i \in \mathcal{I}_{ac}(x^*) \text{ with } \mu_i^* > 0 \ , \\ \nabla g_i(x^*)^T w & \leq & 0 \quad , & i \in \mathcal{I}_{ac}(x^*) \text{ with } \mu_i^* = 0 \ . \end{array} \right.$$

Since $\mu_i^* = 0 \ , \ i \in \mathcal{I}_{in}(x^*)$, we conclude from (2.41)

$$w \in F_2(\lambda^*, \mu^*) \quad \Longleftrightarrow \quad \left\{ \begin{array}{lcll} \lambda_i^* \ \nabla h_i(x^*)^T w & = & 0 \quad , & 1 \leq i \leq m \ , \\ \mu_i^* \ \nabla g_i(x^*)^T w & = & 0 \quad , & 1 \leq i \leq p \ . \end{array} \right. \tag{2.42}$$

Finally, taking the first optimality condition (2.37a) into account

$$\mathcal{L}_x(x^*, \lambda^*, \mu^*) \ = \ \nabla f(x^*) \ + \ \sum_{i=1}^{m} \lambda_i^* \ \nabla h_i(x^*) \ + \ \sum_{i=1}^{p} \mu_i^* \ \nabla g_i(x^*) \ = \ 0 \ ,$$

we find

$$w \in F_2(\lambda^*, \mu^*) \quad \Longrightarrow \tag{2.43}$$
$$\nabla f(x^*)^T w \ = \ - \sum_{i=1}^{m} \lambda_i^* \ \nabla h_i(x^*)^T w \ - \ \sum_{i=1}^{p} \mu_i^* \ \nabla g_i(x^*)^T w \ = \ 0 \ .$$

Consequently, $F_2(\lambda^*, \mu^*)$ contains all those directions from F_1 for which we do not get information from the KKT conditions whether the objective functional f will decrease or increase.

Theorem 2.20 Second order necessary optimality conditions
Assume that $x^* \in \mathcal{F}$ is a local solution of the NLP (2.1a)-(2.1c) and that the LICQ condition holds true. Further, suppose that $\lambda^* \in \mathbb{R}^m$ and $\mu^* \in \mathbb{R}^p$ are Lagrange multipliers satisfying the KKT conditions (2.37a)-(2.37e). Then, the curvature of the Lagrangian is nonnegative along directions in $F_2(\lambda^*, \mu^*)$, i.e.,

$$w^T \ \mathcal{L}_{xx}(x^*, \lambda^*, \mu^*) w \ \geq \ 0 \quad , \quad w \in F_2(\lambda^*, \mu^*) \ . \tag{2.44}$$

Proof: The idea of proof is to construct a feasible sequence $\{x_k\}_{\mathbb{N}} \in \mathcal{T}(x^*)$ with limiting direction $\frac{w}{\|w\|_2}$ such that $f(x_k) \geq f(x^*)$, k sufficiently large, implies (2.44).
Since $w \in F_2(\lambda^*, \mu^*) F_2(\lambda^*, \mu^*) \subset F_1$, for a null sequence $\{t_k\}_{\mathbb{N}}$ of positive real numbers we construct $x_k = z(t_k)$ as in the proof of Lemma 2.13 (ii). In particular, it follows from (2.26a),(2.26b) that

$$h_i(x_k) \ = \ \frac{t_k}{\|w\|_2} \ \nabla h_i(x^*)^T w \quad , \quad 1 \leq i \leq m \ , \tag{2.45a}$$

$$g_i(x_k) \ = \ \frac{t_k}{\|w\|_2} \ \nabla g_i(x^*)^T w \quad , \quad 1 \leq i \leq p \ . \tag{2.45b}$$

Moreover, we have

$$\|x_k \ - \ x^*\|_2 \ = \ t_k \ + \ o(t_k) \ , \tag{2.46}$$

whence

$$x_k \ - \ x^* \ = \ \frac{t_k}{\|w\|_2} \ w \ + \ o(t_k) \ . \tag{2.47}$$

Observing the KKT conditions (2.37a)-(2.37e) and (2.45a),(2.45b), we obtain

$$\begin{aligned}\mathcal{L}(x_k,\lambda^*,\mu^*) \;&=\; f(x_k) \;+\; \sum_{i=1}^{m} \lambda_i^* \; h_i(x_k) \;+\; \sum_{i=1}^{p} \mu_i^* \; g_i(x_k) \;= \\ &=\; f(x_k) \;+\; \frac{t_k}{\|w\|_2} \sum_{i=1}^{m} \underbrace{\lambda_i^* \; \nabla h_i(x^*)^T w}_{=0 \;\text{ due to }\; (2.42)} \;+ \\ &+\; \frac{t_k}{\|w\|_2} \sum_{i=1}^{p} \underbrace{\mu_i^* \; \nabla g_i(x^*)^T w}_{=0 \;\text{ due to }\; (2.42)} \;=\; f(x_k) \;.\end{aligned} \tag{2.48}$$

On the other hand, Taylor expansion yields

$$\begin{aligned}\underbrace{\mathcal{L}(x_k,\lambda^*,\mu^*)}_{=f(x_k) \;\text{ by }\; (2.48)} \;&=\; \underbrace{\mathcal{L}(x^*,\lambda^*,\mu^*)}_{=f(x^*) \;\text{ by }\; \text{KKT}} + (x_k - x^*)^T \; \underbrace{\mathcal{L}_x(x^*,\lambda^*,\mu^*)}_{=0 \;\text{ by }\; \text{KKT}} \\ &+\; \frac{1}{2} \; (x_k - x^*)^T \; \mathcal{L}_{xx}(x^*,\lambda^*,\mu^*)(x_k - x^*) \;+\; o(\|x_k - x^*\|_2^2) \;= \\ &= f(x^*) + \frac{1}{2} \; (x_k - x^*)^T \; \mathcal{L}_{xx}(x^*,\lambda^*,\mu^*)(x_k - x^*) \;+\; o(\|x_k - x^*\|_2^2) \;.\end{aligned} \tag{2.49}$$

Taking (2.46) and (2.47) into account, (2.49) results in

$$f(x_k) \;=\; f(x^*) \;+\; \frac{1}{2} \; \frac{t_k^2}{\|w\|_2^2} \; w^T \mathcal{L}_{xx}(x^*,\lambda^*,\mu^*) w \;+\; o(t_k^2) \;. \tag{2.50}$$

Now, $w^T \mathcal{L}_{xx}(x^*,\lambda^*,\mu^*) w < 0$ would imply that $f(x_k) < f(x^*)$ for sufficiently large $k \in \mathbb{N}$, which contradicts the assumption that x^* is a local solution of the NLP (2.1a)-(2.1c). Consequnetly, (2.44) must hold true.

•

If we require $\mathcal{L}_{xx}(x^*,\lambda^*,\mu^*)$ to be uniformly positive definite on $F_2(\lambda^*,\mu^*)$, then this constitutes a sufficient condition for optimality. Note that the LICQ condition is not required.

Theorem 2.21 Second order sufficient optimality conditions

Assume that $x^* \in \mathcal{F}$ is a feasible point and there exist Lagrange multipliers $\lambda^* \in \mathbb{R}^m$ and $\mu^* \in \mathbb{R}^p$ satisfying the KKT conditions (2.37a)-(2.37e). Further, suppose that

$$w^T \mathcal{L}_{xx}(x^*,\lambda^*,\mu^*) w \;>\; 0 \quad , \quad w \in F_2(\lambda^*,\mu^*) \;. \tag{2.51}$$

Then, x^* is a strict local solution of the NLP (2.1a)-(2.1c).

Proof: We will show that for any feasible sequence $\{x_k\}_{\mathbb{N}} \in \mathcal{T}(x^*)$ with $x_k \to x^*$ $(k \to \infty)$ we have that

$$f(x_k) \;>\; f(x^*) \quad \text{for sufficiently large } k \in \mathbb{N} \;. \tag{2.52}$$

By Lemma 2.13 (i) and Definition 2.14 we have that all limiting directions d of $\{x_k\}_{\mathbb{N}}$ satisfy $d \in F_1$. Let d be a limiting direction and $N' \subset \mathbb{N}$ such that for all $k \in \mathbb{N}'$

$$x_k \;-\; x^* \;=\; \|x_k \;-\; x^*\|_2 \, d \;+\; o(\|x_k \;-\; x^*\|_2) \;. \tag{2.53}$$

For the Lagrangian we have

$$\mathcal{L}(x_k,\lambda^*,\mu^*) \;=\; f(x_k) \;+\; \sum_{i\in\mathcal{I}_{ac}(x^*)} \mu_i^* g_i(x_k) \;\le\; f(x_k) \;. \tag{2.54}$$

We will prove (2.52) for the two cases $d \notin F_2(\lambda^*,\mu^*)$ and $d \in F_2(\lambda^*,\mu^*)$.

Case I: $d \notin F_2(\lambda^*,\mu^*)$

In view of (2.41) there exists an index $j \in \mathcal{I}_{ac}(x^*)$ such that

$$\mu_j^* \nabla g_j(x^*)^T d \;<\; 0 \;, \tag{2.55a}$$

$$\mu_i^* \nabla g_i(x^*)^T d \;<\; 0 \;,\; i \in \mathcal{I}_{ac}(x^*) \setminus \{j\} \;. \tag{2.55b}$$

Observing (2.53), by Taylor expansion we get for $k \in \mathbb{N}'$

$$\begin{aligned} \mu_j^* g_j(x_k) \;&=\; \underbrace{\mu_j^* g_j(x^*)}_{=\,0} \;+\mu_j^* \nabla g_j(x^*)^T (x_k - x^*) \;+\; o(\|x_k - x^*\|_2) \;= \\ &=\; \|x_k - x^*\|_2 \; \mu_j^* \nabla g_j(x^*)^T d \;+\; o(\|x_k - x^*\|_2) \;. \end{aligned}$$

Consequently, (2.54) infers

$$\begin{aligned} \mathcal{L}(x_k,\lambda^*,\mu^*) \;&=\; f(x_k) \;+\; \sum_{i\in\mathcal{I}_{ac}(x^*)} \mu_i^* g_i(x_k) \;\le \\ &\le\; f(x_k) \;+\; \|x_k - x^*\|_2 \; \mu_j^* \nabla g_j(x^*)^T d \;+\; o(\|x_k - x^*\|_2) \;. \end{aligned} \tag{2.56}$$

On the other hand, using the Taylor expansion (2.49) in the proof of Theorem 2.20, we have

$$\mathcal{L}(x_k,\lambda^*,\mu^*) \;=\; f(x^*) \;+\; O(\|x_k - x^*\|_2^2) \;. \tag{2.57}$$

Combining (2.56) and (2.57) yields

$$f(x_k) \;\ge\; f(x^*) \;-\; \|x_k - x^*\|_2 \; \mu_j^* \nabla g_j(x^*)^T d \;+\; o(\|x_k - x^*\|_2) \;.$$

Observing (2.55a), we deduce (2.52).

Case II: $d \in F_2(\lambda^*,\mu^*)$

In this case, the Taylor expansion (2.49) and (2.53),(2.54) imply

$$\begin{aligned} f(x_k) \;&\ge\; f(x^*) \;+\; \frac{1}{2}\,(x_k - x^*)^T \mathcal{L}_{xx}(x^*,\lambda^*,\mu^*)(x_k - x^*) \;+\; o(\|x_k - x^*\|_2^2) \;= \\ &=\; f(x^*) \;+\; \frac{1}{2}\,\|x_k - x^*\|_2^2 \; \underbrace{d^T \mathcal{L}_{xx}(x^*,\lambda^*,\mu^*) d}_{>\;0\;\;\text{by}\;\;(2.51)} \;+ o(\|x_k - x^*\|_2^2) \;, \end{aligned}$$

from which we deduce (2.52).
Finally, since each x_k , $k \in \mathbb{N}$, can be assigned to one of the subsequences $\mathbb{N}' \subset \mathbb{N}$ such that $\{x_k\}_{\mathbb{N}'}$ converges to a limiting direction d, we have that $f(x_k) > f(x^*)$ for sufficiently large $k \in \mathbb{N}$.

•

Example 2.22 Second order optimality conditions
We reconsider Example 2.6 where the Lagrangian is given by

$$\mathcal{L}(x,\lambda) \ = \ (x_1 + x_2) \ + \ \lambda \ (x_1^2 + x_2^2 - 2) \ .$$

The KKT conditions (2.37a)-(2.37e) are satisfied for $x^* = (-1,-1)^T$ and $\lambda^* = 0.5$. The Hessian of the Lagrangian at (x^*,λ^*) is given by

$$\mathcal{L}_{xx}(x^*,\lambda^*) \ = \ \begin{pmatrix} 2\lambda^* & 0 \\ 0 & 2\lambda^* \end{pmatrix} \ = \ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \ .$$

Obviously, $w^T\mathcal{L}_{xx}(x^*,\lambda^*)w > 0$ for all $w \neq 0$, and hence, it follows from Theorem 2.21 that x^* is a strict local solution.

Example 2.23 Second order optimality conditions
Consider the inequality constrained NLP

$$\begin{aligned} &\text{minimize} \quad f(x) := -0.1(x_1-4)^2 + x_2^2 && (2.58a)\\ &\text{over} \quad x \in \mathbb{R}^2 \\ &\text{subject to} \quad g(x) := 1 - (x_1^2 + x_2^2) \le 0 \ , && (2.58b) \end{aligned}$$

Obviously, the objective functional f is not bounded from below on the feasible set $\mathcal{F}$, and hence, no global solution exists.
For the associated Lagrangian we obtain

$$\mathcal{L}_x(x,\mu) \ = \ \begin{pmatrix} -0.2(x_1-4) \ - \ 2\mu x_1 \\ 2x_2 \ - \ 2\mu x_2 \end{pmatrix} \ , \qquad (2.59a)$$

$$\mathcal{L}_{xx}(x,\mu) \ = \ \begin{pmatrix} -0.2 - 2\mu & 0 \\ 0 & 2 - 2\mu \end{pmatrix} \ . \qquad (2.59b)$$

The pair (x^*,μ^*) with $x^* = (1,0)^T$ and $\mu^* = 0.3$ satisfies the KKT conditions (2.37a)-(2.37e). In order to check the second order sufficient optimality condition (2.51), we note that

$$\nabla g(x^*) \ = \ \begin{pmatrix} -2x_1^* \\ -2x_2^* \end{pmatrix} \ = \ \begin{pmatrix} -2 \\ 0 \end{pmatrix} \ ,$$

whence

$$F_2(\mu^*) \ = \ \{ \ w = (w_1,w_2)^T \in \mathbb{R}^2 \ | \ w_1 = 0 \ \} \ .$$

Consequently, it follows from (2.59b) that for $w \in F_2(\mu^*)$

$$w^T \mathcal{L}_{xx}(x^*, \mu^*) w \;=\; \begin{pmatrix} 0 \\ w_2 \end{pmatrix}^T \begin{pmatrix} -0.8 & 0 \\ 0 & 1.4 \end{pmatrix} \begin{pmatrix} 0 \\ w_2 \end{pmatrix} \;=\; 1.4 w_2^2 \;>\; 0 \;.$$

Hence, (2.51) is satisfied and thus, x^* is a strict local solution for (2.58a),(2.58b).

2.4 Projected Hessians

The second order conditions (2.44) and (2.51) are now stated in a weaker form that can be checked numerically in an easier way. We assume that the triple (x^*, λ^*, μ^*) with uniquely determined Lagrange multipliers λ^* , μ^* fulfills the KKT conditions (2.37a)-(2.37e) and distinguish the two cases where either strict complementarity (2.39) is satisfied or does not hold true.

Case I: KKT & Strict complementarity

We assume that the KKT conditions (2.37a)-(2.37e) are fulfilled and strict complementarity (2.39) holds true. In this case, the definition of $F_2(\lambda^*, \mu^*)$ (cf. (2.42)) reduces to

$$F_2(\lambda^*, \mu^*) \;=\; \mathrm{Ker}\; A \;,$$

where the matrix A is given by (2.27). It follows that the matrix Z defined by (2.28) has full column rank (since λ^*, μ^* are uniquely determined), and its columns span the space $F_2(\lambda^*, \mu^*)$. Hence, $w \in F_2(\lambda^*, \mu^*)$ if and only if $w = Zu$ for some $u \in \mathbb{R}^{n-q^*}$. Consequently, the conditions (2.44) in Theorem 2.20 and (2.51) in Theorem 2.21 can be written as follows

$$u^T Z^T \mathcal{L}_{xx}(x^*, \lambda^*, \mu^*) Z u \;\geq\; 0 \quad , \quad u \in \mathbb{R}^{n-q^*} \qquad , \tag{2.60a}$$

$$u^T Z^T \mathcal{L}_{xx}(x^*, \lambda^*, \mu^*) Z u \;>\; 0 \quad , \quad u \in \mathbb{R}^{n-q^*} \setminus \{0\} \;, \tag{2.60b}$$

which means that $Z^T \mathcal{L}_{xx}(x^*, \lambda^*, \mu^*) Z$ is positive semidefinite and positive definite, respectively.

Case II: KKT without strict complementarity

We suppose that the KKT conditions (2.37a)-(2.37e) with uniquely determined Lagrange multipliers λ^*, μ^* are satisfied, but do not assume strict complementarity (2.39). In this case, $F_2(\lambda^*, \mu^*)$ is not a subspace but the intersection of the planes defined by the first two conditions in (2.41) and the half-spaces defined by the third condition in (2.41).

We introduce $\underline{F_2}$ and $\overline{F}_2$ according to

$$\underline{F}_2 := \{d \in F_1 \mid \left\{ \begin{array}{lcl} \nabla h_i(x^*)^T d = 0 & , & 1 \leq i \leq m \\ \nabla g_i(x^*)^T d = 0 & , & i \in \mathcal{I}_{ac}(x^*) \end{array} \right\}\} \;, \tag{2.61a}$$

$$\overline{F}_2 := \{d \in F_1 \mid \left\{ \begin{array}{lcl} \nabla h_i(x^*)^T d = 0 & , & 1 \leq i \leq m \\ \nabla g_i(x^*)^T d = 0 & , & i \in \mathcal{I}_{ac}(x^*) \text{ or } \mu_i^* > 0 \end{array} \right\}\} \;. \tag{2.61b}$$

Note that $\underline{F_2}$ is the largest-dimensional subspace that is contained in $F_2(\lambda^*, \mu^*)$, whereas $\overline{F}_2$ is the smallest-dimensional subspace that contains $F_2(\lambda^*, \mu^*)$.
As in case I we construct matrices $\underline{A}$ and $\overline{A}$ as well as $\underline{Z}$ and $\overline{Z}$ whose columns span Ker $\underline{A} = \underline{F_2}$ and Ker $\overline{A} = \overline{F}_2$.
Now, if (2.44) in Theorem 2.20 is satisfied, due to $\underline{F_2} \subset F_2(\lambda^*, \mu^*)$ we have

$$w^T \mathcal{L}_{xx}(x^*, \lambda^*, \mu^*) w \;\geq\; 0 \quad , \quad w \in \underline{F_2} \ ,$$

i.e., $\underline{Z}^T \mathcal{L}_{xx}(x^*, \lambda^*, \mu^*) \underline{Z}$ is positive semidefinite.
Likewise, the condition

$$w^T \mathcal{L}_{xx}(x^*, \lambda^*, \mu^*) w \;>\; 0 \quad , \quad w \in \overline{F}_2 \ ,$$

i.e., $\overline{Z}^T \mathcal{L}_{xx}(x^*, \lambda^*, \mu^*) \overline{Z}$ is positive definite, implies (2.52) in Theorem 2.21.

Definition 2.24 Projected Hessians

The matrices $Z^T \mathcal{L}_{xx}(x^*, \lambda^*, \mu^*) Z$ and $\underline{Z}^T \mathcal{L}_{xx}(x^*, \lambda^*, \mu^*) \underline{Z}$, $\overline{Z}^T \mathcal{L}_{xx}(x^*, \lambda^*, \mu^*) \overline{Z}$ are called projected Hessians.

The matrices Z and $\underline{Z}, \overline{Z}$ can be computed by a QR-factorization of the matrices A^T and $\underline{A}, \overline{A}$. If $A \in \mathbb{R}^{q^* \times n}$ has full row rank q^*, we obtain

$$A^T \;=\; Q \begin{bmatrix} R \\ 0 \end{bmatrix} \;=\; [Q_1 \; Q_2] \begin{bmatrix} R \\ 0 \end{bmatrix} \;=\; Q_1 R \ , \qquad (2.62)$$

where $R \in \mathbb{R}^{q^* \times q^*}$ is a regular upper triangular matrix and $Q \in \mathbb{R}^{n \times n}$ is orthogonal. Moreover, $Q_1 \in \mathbb{R}^{n \times q^*}$, $Q_2 \in \mathbb{R}^{n \times (n - q^*)}$. Since

$$AQ \;=\; [R^T \; 0] \; Q^T Q \;=\; [R^T \; 0]$$

and R is nonsingular, we find $Z = Q_2$.
If A has row rank $\hat{q} < q^*$, we perform column pivoting during the QR-factorization of A^T. This means that we get a QR-factorization

$$A^T P \;=\; Q \begin{bmatrix} R \\ 0 \end{bmatrix} \;=\; [Q_1 \; Q_2] \begin{bmatrix} R \\ 0 \end{bmatrix} \;=\; Q_1 R \ , \qquad (2.63)$$

where $P \in \mathbb{R}^{q^* \times n}$ is a permutation matrix, $R \in \mathbb{R}^{\hat{q} \times \hat{q}}$ is upper triangular and regular, and $Q_1 \in \mathbb{R}^{n \times \hat{q}}$, $Q_2 \in \mathbb{R}^{n \times (n - \hat{q})}$ have orthonormal columns. Again, we obtain $Z = Q_2$.