# Chapter 4 Sequential Quadratic Programming

## 4.1 The Basic SQP Method

### 4.1.1 Introductory Definitions and Assumptions

Sequential Quadratic Programming (SQP) is one of the most successful methods for the numerical solution of constrained nonlinear optimization problems. It relies on a profound theoretical foundation and provides powerful algorithmic tools for the solution of large-scale technologically relevant problems.

We consider the application of the SQP methodology to nonlinear optimization problems (NLP) of the form

$$\begin{aligned} &\text{minimize} \quad f(x) && (4.1a)\\ &\text{over} \quad x \in \mathbb{R}^n && \\ &\text{subject to} \quad h(x) = 0 && (4.1b)\\ &\qquad\qquad\quad g(x) \le 0 \ , && (4.1c) \end{aligned}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is the objective functional, 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.

The NLP (4.1a)-(4.1c) contains as special cases linear and quadratic programming problems, when $f$ is linear or quadratic and the constraint functions $h$ and $g$ are affine.

SQP is an iterative procedure which models the NLP for a given iterate $x^k$ , $k \in \mathbb{N}_0$, by a Quadratic Programming (QP) subproblem, solves that QP subproblem, and then uses the solution to construct a new iterate $x^{k+1}$. This construction is done in such a way that the sequence $(x^k)_{k\in\mathbb{N}_0}$ converges to a local minimum $x^*$ of the NLP (4.1a)-(4.1c) as $k \to \infty$. In this sense, the NLP resembles the Newton and quasi-Newton methods for the numerical solution of nonlinear algebraic systems of equations. However, the presence of constraints renders both the analysis and the implementation of SQP methods much more complicated.

**Definition 4.1 Feasible set**

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

$$\mathcal{F} \ := \ \{ \ x \in \mathbb{R}^n \mid h(x) \ = \ 0 \ , \ g(x) \ \le \ 0 \ \} \qquad (4.2)$$

is called the feasible set of the NLP (4.1a)-(4.1c). Its elements are referred to as feasible points.

Note that a major advantage of SQP is that the iterates $x^k$ need not to be feasible points, since the computation of feasible points in case of nonlinear constraint functions may be as difficult as the solution of the NLP itself.

**Definition 4.2 Lagrangian functional associated with the NLP**
The functional $\mathcal{L} : \mathbb{R}^{n \times m \times p} \to \mathbb{R}$ defined by means of

$$\mathcal{L}(x, \lambda, \mu) \ := \ f(x) \ + \ \lambda^T h(x) \ + \ \mu^T g(x) \tag{4.3}$$

is called the Lagrangian functional of the NLP (4.1a)-(4.1c). The vectors $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^p_+$ are referred to as Lagrangian multipliers.

For a functional $f : \mathbb{R}^n \to \mathbb{R}$, we denote by $\nabla f(x)$ the gradient of $f$ at $x \in \mathbb{R}^n$, i.e.,

$$\nabla f(x) \ := \ \Big(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, ..., \frac{\partial f(x)}{\partial x_n}\Big)^T \ .$$

We further refer to $Hf(x)$ as the Hessian of $f$ at $x \in \mathbb{R}^n$, i.e., the matrix of second partial derivatives as given by

$$(Hf(x))_{ij} \ := \ \frac{\partial^2 f(x)}{\partial x_i \partial x_j} \quad , \quad 1 \le i, j \le n \ .$$

For vector-valued functions $h : \mathbb{R}^n \to \mathbb{R}^m$ the symbol $\nabla$ is also used for the Jacobian of $h$ according to

$$\nabla h(x) \ := \ \big(\nabla h_1(x), \nabla h_2(x), ..., \nabla h_m(x)\big) \ .$$

**Definition 4.3 Set of active constraints**
For $x \in \mathbb{R}^n$, the index set

$$\mathcal{I}_{ac}(x) \ := \ \{ \ i \in \{1, ..., p\} \mid g_i(x) \ = \ 0 \ \} \tag{4.4}$$

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

**Definition 4.4 Strict complementary slackness**
If $x^* \in \mathbb{R}^n$ is a local minimum of the NLP (4.1a)-(4.1c), the condition

$$g_i(x^*)\mu_i^* \ = \ 0 \quad , \quad 1 \le i \le p \ , \tag{4.5}$$
$$\mu_i^* \ > \ 0 \quad , \quad i \in \mathcal{I}_{ac}(x^*) \tag{4.6}$$

is called strict complementary slackness at $x^*$.

Setting $q_x := |\mathcal{I}_{ac}(x)|$ and assuming $\mathcal{I}_{ac}(x) = \{i_1, ..., i_{q(x)}\}$, we will denote by $G(x) \in \mathbb{R}^{n \times (m+q_x)}$ the matrix given by

$$G(x) \ := \ \big(\nabla h_1(x), \nabla h_2(x), ..., \nabla h_m(x), \nabla g_{i_1}(x), ..., \nabla g_{i_{q_x}}(x)\big) \ .$$

Throughout the following, we suppose that the functions $f, g$, and $h$ are three times continuously differentiable.

**Definition 4.5 First order necessary optimality conditions**
Let $x^* \in \mathbb{R}^n$ be a local minimum of the NLP (4.1a)-(4.1c) and suppose there exist Lagrange multipliers $\lambda^* \in \mathbb{R}^m$ and $\mu^* \in \mathbb{R}^p_+$ such that

$$(\mathbf{A1}) \qquad \nabla \mathcal{L}(x^*, \lambda^*, \mu^*) \ = \ \nabla f(x^*) \ + \ \nabla h(x^*)\lambda^* \ + \ \nabla g(x^*)\mu^* \ = \ 0$$

holds true. Then, ($\mathbf{A1}$) is referred to as the first order necessary optimality or Karush-Kuhn-Tucker (KKT) conditions.

**Definition 4.6 Critical points**
A feasible point $x \in \mathcal{F}$ that satisfies the first order necessary optimality conditions ($\mathbf{A1}$) is called a critical point of the NLP (4.1a)-(4.1c).

Note that a critical point need not to be a local minimum.

**Definition 4.7 Second order sufficient optimality conditions**
In addition to ($\mathbf{A1}$) suppose that the following conditions are satisfied:

($\mathbf{A2}$) The columns of $G(x^*)$ are linearly independent.

($\mathbf{A3}$) Strict complementary slackness holds at $x^*$.

($\mathbf{A4}$) The Hessian of the Lagrangian with respect to $x$ is positive definite on the null space of $G(x^*)^T$, i.e.,

$$d^T H\mathcal{L}^* d \ > \ 0 \quad \text{for all } d \neq 0 \text{ such that } G(x^*)^T d = 0 \ .$$

The conditions ($\mathbf{A1}$)-($\mathbf{A4}$) are called the second order sufficient optimality conditions of the NLP (4.1a)-(4.1c).

The second order sufficient optimality conditions guarantee that $x^*$ is an isolated local minimum of the NLP (4.1a)-(4.1c) and that the Lagrange multipliers $\lambda^*$ and $\mu^*$ are uniquely determined.

The convergence behavior of SQP methods will be measured by asymptotic convergence rates with respect to the Euclidean norm $\|\cdot\|$.

**Definition 4.8 Convergence rates**
Let $(x^k)_{k\in\mathbb{N}_0}$ be a sequence of iterates converging to $x^*$. The sequence is said to convergence

- linearly, if there exist $0 < q < 1$ and $k_{max} \geq 0$ such that for all $k \geq k_{max}$

$$\|x^{k+1} \ - \ x^*\| \ \leq \ q \ \|x^k \ - \ x^*\| \ ,$$

- superlinearly, if there exist a null sequence $(q_k)_{k\in\mathbb{N}_0}$ of positive numbers and $k_{max} \geq 0$ such that for all $k \geq k_{max}$

$$\|x^{k+1} \ - \ x^*\| \ \leq \ q_k \ \|x^k \ - \ x^*\| \ ,$$

- quadratically, if there exist $c > 0$ and $k_{max} \geq 0$ such that for all $k \geq k_{max}$

$$\|x^{k+1} - x^*\| \leq c \, \|x^k - x^*\|^2 .$$

- R-linearly, if there exist $0 < q < 1$ such that

$$\limsup_{k\to\infty} \sqrt[k]{\|x^k - x^*\|} \leq \sqrt[k]{q} .$$

### 4.1.2 Construction of the QP Subproblems

The QP subproblems which have to be solved in each iteration step should reflect the local properties of the NLP with respect to the current iterate $x^k$. Therefore, a natural idea is to replace the

- objective functional $f$ by its local quadratic approximation

$$f(x) \approx f(x^k) + \nabla f(x^k)(x - x^k) + \frac{1}{2} (x - x^k)^T Hf(x^k)(x - x^k) ,$$

- constraint functions $g$ and $h$ by their local affine approximations

$$\begin{aligned} g(x) &\approx g(x^k) + \nabla g(x^k)(x - x^k) , \\ h(x) &\approx h(x^k) + \nabla h(x^k)(x - x^k) . \end{aligned}$$

Setting

$$d(x) := x - x^k \quad , \quad B_k := Hf(x^k) , \tag{4.7}$$

this leads to the following form of the QP subproblem:

$$\text{minimize} \quad \nabla f(x^k)^T d(x) + \frac{1}{2} d(x)^T B_k d(x) \tag{4.8a}$$

$$\text{over} \quad d(x) \in \mathbb{R}^n$$

$$\text{subject to} \quad h(x^k) + \nabla h(x^k)^T d(x) = 0 \tag{4.8b}$$

$$g(x^k) + \nabla g(x^k)^T d(x) \leq 0 , \tag{4.8c}$$

**Remark 4.9** The following example shows that the computation of the increment $d(x)$ as the solution of the associated QP may break down. Consider the NLP

$$\begin{aligned} &\text{minimize} \quad - x_1 - \frac{1}{2} x_2^2 \\ &\text{over} \quad x = (x_1, x_2)^T \in \mathbb{R}^2 \\ &\text{subject to} \quad x_1^2 + x_2^2 - 1 = 0 . \end{aligned}$$

Obviously, $x^* = (1,0)^T$ is a solution satisfying the second order sufficient optimality conditions (**A1**)-(**A4**).
Choosing $x^k = (1+\varepsilon, 0)^T$ , $\varepsilon > 0$, and observing

$$\nabla f(x) = \begin{pmatrix} -1 \\ -x_2 \end{pmatrix} \quad , \quad Hf(x) = \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix} \; ,$$
$$\nabla h(x) = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix} \; ,$$

the QP takes the form

$$\begin{aligned} &\text{minimize} \quad -d_1(x) - \frac{1}{2} d_2(x)^2 \\ &\text{over} \quad d(x) = (d_1(x), d_2(x))^T \in \mathbb{R}^2 \\ &\text{subject to} \quad d_1(x) = -\frac{1}{2}\, \varepsilon\, \frac{2+\varepsilon}{1+\varepsilon} \; . \end{aligned}$$

Clearly, the QP is unbounded no matter how small $\varepsilon$ is chosen.
With regard to convergence conditions for the SQP to be discussed in section 2.3, we remark that the Hessian $Hf$ of $f$ is singular in this particular example.

The QP (4.8a)-(4.8c) is related to a local quadratic model of the Lagrangian $\mathcal{L}$ as the objective functional which leads to the QP subproblem

$$\text{minimize} \quad \nabla \mathcal{L}(x^k, \lambda^k, \mu^k)^T d(x) + \frac{1}{2} d(x)^T H\mathcal{L}(x^k, \lambda^k, \mu^k) d(x) \tag{4.9a}$$
$$\text{over} \quad d(x) \in \mathbb{R}^n$$
$$\text{subject to} \quad h(x^k) + \nabla h(x^k)^T d(x) = 0 \tag{4.9b}$$
$$g(x^k) + \nabla g(x^k)^T d(x) \le 0 \; , \tag{4.9c}$$

where $\lambda^k$ and $\mu^k$ are the Lagrangian multipliers associated with this QP.
Considering the QP (4.9a)-(4.9c) is justified by the fact that the conditions (**A1**)-(**A4**) imply that $x^*$ is a local minimum of the problem

$$\begin{aligned} &\text{minimize} \quad \mathcal{L}(x, \lambda^*, \mu^*) \\ &\text{over} \quad x \in \mathbb{R}^n \\ &\text{subject to} \quad h(x) = 0 \\ &\qquad\qquad\quad\; g(x) \le 0 \; . \end{aligned}$$

Indeed, given an iterate $(x^k, \lambda^k, \mu^k)$, the objective functional in (4.9a) stems from the quadratic Taylor series approximation

$$\mathcal{L}(x^k, \lambda^k, \mu^k) + \nabla \mathcal{L}(x^k, \lambda^k, \mu^k)^T d(x) + \frac{1}{2} d(x)^T H\mathcal{L}(x^k, \lambda^k, \mu^k) d(x)$$

of the Lagrangian $\mathcal{L}$ in $x$.

In general, the two QP subproblems (4.8a)-(4.8c) and (4.9a)-(4.9c) are not equivalent. However, in special cases their equivalence can be established.

**Lemma 4.10 Equivalence of QP subproblems, Part I**

Consider the NLP (4.1a)-(4.1c) and the two QP subproblems (4.8a)-(4.8c),(4.9a)-(4.9c). Then, there holds:

(i) If there are no inequality constraints in the NLP (4.1a)-(4.1c), then the QP subproblems (4.8a)-(4.8c) and (4.9a)-(4.9c) are equivalent.

(ii) In the fully constrained case, assume that the multiplier $\mu^k$ in (4.9a)-(4.9c) satisfies

$$\mu_i^k \; = \; 0 \quad , \quad i \in \mathcal{I}_{in}(x^k) \; . \tag{4.10}$$

Then, the SQ subproblems (4.8a)-(4.8c) and (4.9a)-(4.9c) are equivalent.

**Proof:** For the proof of part (i) we observe that due to the linearized equality constraints the term $\nabla h(x^k)^T d(x)$ is a constant. Hence, in view of (4.3), the objective functional in (4.9a) reduces to that in (4.8a).
The proof of part (ii) can be done in the same way taking into account that in view of (4.10) only the active inequality constraints come into play. Since $\nabla h(x^k)^T d(x)$ and $\nabla g_i(x^k)^T d(x), i \in \mathcal{I}_a(x^k)$ are constants, we conclude.

•

The second part of the previous lemma motivates to consider the slack-variable formulation of the NLP (4.1a)-(4.1c) which can be stated in terms of an additional vector $z \in \mathbb{R}^p$ called the vector of slack variables:

$$\begin{aligned} &\text{minimize} \quad f(x) && (4.11a)\\ &\text{over} \quad (x,z) \in \mathbb{R}^n \times \mathbb{R}^p_+ \\ &\text{subject to} \qquad h(x) = 0 && (4.11b)\\ &\qquad\qquad g(x) + z \; = \; 0 \; . && (4.11c)\end{aligned}$$

For the slack-variable formulation (4.11a)-(4.11c) of the NLP (4.1a)-(4.1c) we can show the equivalence of the associated QP subproblems.

**Lemma 4.11 Equivalence of QP subproblems, Part II**

Consider the QP subproblems associated with the slack-variable formulation (4.11a)-(4.11c) of the NLP (4.1a)-(4.1c) as given by

$$\begin{aligned} &\text{minimize} \quad \nabla f(x^k)^T d(x) \; + \; \frac{1}{2} d(x)^T B_k d(x) && (4.12a)\\ &\text{over} \quad (d(x), d(z)) \in \mathbb{R}^n \times \mathbb{R}^p_+ \\ &\text{subject to} \qquad h(x^k) \; + \; \nabla h(x^k)^T d(x) \; = \; 0 && (4.12b)\\ &\qquad\qquad g(x^k) \; + \; \nabla g(x^k)^T d(x) \; + \; d(z) \; = \; 0 && (4.12c)\end{aligned}$$

and

$$\text{minimize} \quad \nabla \mathcal{L}(x^k, \lambda^k, \mu^k)^T d(x) \ + \ \frac{1}{2} d(x)^T H\mathcal{L}(x^k, \lambda^k, \mu^k) d(x) \tag{4.13a}$$

$$\text{over} \quad (d(x), d(z)) \in \mathbb{R}^n \times \mathbb{R}^p_+$$

$$\text{subject to} \qquad h(x^k) \ + \ \nabla h(x^k)^T d(x) \ = \ 0 \tag{4.13b}$$

$$g(x^k) \ + \ \nabla g(x^k)^T d(x) \ + \ d(z) \ = 0 \ . \tag{4.13c}$$

The two QP subproblems (4.12a)-(4.12c) and (4.13a)-(4.13c) are equivalent.

**Proof:** The proof is left as an exercise.

•

### 4.2 Local convergence

The local convergence analysis for the SQP method will be carried out under the assumption that the active inequality constraints of the NLP at the local minimum $x^*$ are known which is justified by the fact that the QP subproblem for the iterate $x^k$ has the same active constraints at $x^k$, provided $x^k$ is sufficiently close to $x^*$.
Consequently, we restrict the analysis to QP problems of the form:

$$\text{minimize} \quad \nabla f(x^k)^T d_x \ + \ \frac{1}{2} \ d_x^T B_k d_x \ , \tag{4.14a}$$

$$\text{subject to} \quad \nabla h(x^k)^T d_x \ + \ h(x^k) \ = \ 0 \ . \tag{4.14b}$$

**Remark 4.12 Starting value for the Lagrange multiplier**
A good starting value $x^0$ for the solution can be used to obtain an appropriate starting value $\lambda^0$ for the Lagrange multiplier. Indeed, the first order necessary condition **(A1)** and condition **(A2)** imply

$$\lambda^* \ = \ - \ \left[\nabla h(x^*)^T A \nabla h(x^*)\right]^{-1} \ \nabla h(x^*)^T A \nabla f(x^*) \tag{4.15}$$

for any nonsingular $A$ which is positive definite on the null space of $\nabla h(x^*)^T$. In particular, if $A$ is chosen as the identity matrix, then (4.15) defines the least squares solution of the first order necessary optimality conditions. Consequently,

$$\lambda^0 \ = \ - \ \left[\nabla h(x^0)^T \nabla h(x^0)\right]^{-1} \ \nabla h(x^0)^T \nabla f(x^0) \tag{4.16}$$

is close to $\lambda^*$, if $x^0$ is close to $x^*$.

**Remark 4.13 KKT conditions for the quadratic subproblem**
Denoting the optimal multiplier for (4.14a)-(4.14b) by $\lambda^{k+1}$, the KKT conditions are as follows:

$$\begin{aligned} B_k d_x \ + \ \nabla h(x^k) \lambda^{k+1} \ &= \ - \ \nabla f(x^k) \ , \\ \nabla h(x^k)^T d_x \ &= \ - \ h(x^k) \ . \end{aligned}$$

Thus, setting

$$d_\lambda \ := \ \lambda^{k+1} \ - \ \lambda^k \ , \tag{4.17}$$

they can be rewritten according to

$$B_k d_x \ + \ \nabla h(x^k) d_\lambda \ = \ - \ \nabla \mathcal{L}(x^k, \lambda^k) \ , \tag{4.18}$$

$$\nabla h(x^k)^T d_x \ = \ - \ h(x^k) \ . \tag{4.19}$$

### 4.2.1 The Newton SQP method

The local convergence of the SQP method follows from the application of Newton's method to the nonlinear system given by the KKT conditions

$$\Psi(x, \lambda) \ = \ \begin{bmatrix} \nabla \mathcal{L}(x, \lambda) \\ h(x) \end{bmatrix} \ = \ 0 \ .$$

Assumptions **(A2)** and **(A4)** imply that the Jacobian

$$J(x^*, \lambda^*) \ = \ \begin{bmatrix} H\mathcal{L}(x^*, \lambda^*) & \nabla h(x^*) \\ \nabla h(x^*)^T & 0 \end{bmatrix} \tag{4.20}$$

at the local solution $(x^*, \lambda^*)$ is nonsingular.
Therefore, the Newton iteration

$$x^{k+1} \ = \ x^k \ + \ s_x \ ,$$
$$\lambda^{k+1} \ = \ \lambda^k \ + \ s_\lambda \ ,$$

where $s = (s_x, s_\lambda)$ is the solution of

$$J(x^k, \lambda^k) s \ = \ - \ \Psi(x^k, \lambda^k) \ , \tag{4.21}$$

converges quadratically, provided $(x^0, \lambda^0)$ is sufficiently close to $(x^*, \lambda^*)$.
The equations (4.21) correspond to (4.18),(4.19) for $B_k = H\mathcal{L}(x^k, \lambda^k), d_x = s_x$, and $d_\lambda = s_\lambda$.
Consequently, the iterates $(x^{k+1}, \lambda^{k+1})$ are exactly those generated by the SQP algorithm.
We have thus proved:

**Theorem 4.14 Convergence of the Newton SQP method**

Let $x^0$ be a starting value for the solution of the NLP. Assume the starting value $\lambda^0$ to be given by (4.16). Suppose further that the sequence of iterates is given by

$$x^{k+1} \ = \ x^k \ + \ d_x \ ,$$
$$\lambda^{k+1} \ = \ \lambda^k \ + \ d_\lambda \ ,$$

where $d_x$ and $\lambda^k + d_\lambda$ are the solution and the associated multiplier of the QP subproblem (4.14a),(4.14b) with $B_k = H\mathcal{L}(x^k, \lambda^k)$.
If $\|x^0 - x^*\|$ is sufficiently small, the sequence of iterates is well defined and converges quadratically to the optimal solution $(x^*, \lambda^*)$.

### 4.2.2 Conditions for local convergence

We impose the following conditions on $H\mathcal{L}(x^*, \lambda^*)$ and $B_k$ that guarantee local convergence of the SQP algorithm:

**(A5)** The matrix $H\mathcal{L}(x^*, \lambda^*)$ is nonsingular.

**(B1)** The matrices $B_k$ are uniformly positive definite on the null spaces of $\nabla h(x^k)^T$, i.e., there exists $\beta_1 > 0$ such that for all $k$

$$d^T B_k d \;\geq\; \beta_1 \|d\|^2 \quad \text{for all } d \text{ such that } \nabla h(x^k)^T d = 0 \; .$$

**(B2)** The sequence $\{B_k\}_{\mathbb{N}}$ is uniformly bounded, i.e., there exists $\beta_2 > 0$ such that for all $k$

$$\|B_k\| \;\leq\; \beta_2 \; .$$

**(B3)** The matrices $B_k$ have uniformly bounded inverses, i.e., there is a constant $\beta_3 > 0$ such that $B_k^{-1}$ exists and

$$\|B_k^{-1}\| \;\leq\; \beta_3 \; .$$

**Theorem 4.15 Linear convergence of the SQP algorithm**
Assume that the conditions **(A1)-(A5)** and **(B1)-(B3)** hold true. Let $\mathcal{P}$ be the projection operator

$$\mathcal{P}(x) \;:=\; I \;-\; \nabla h(x) \left[\nabla h(x)^T \nabla h(x)\right]^{-1} \nabla h(x)^T \; . \qquad (4.22)$$

Then there exist constants $\varepsilon > 0$ and $\gamma > 0$ such that if

$$\|x^0 \;-\; x^*\| \;<\; \varepsilon \quad , \quad \|\lambda^0 \;-\; \lambda^*\| \;<\; \varepsilon$$

and

$$\|\mathcal{P}(x^k)(B_k \;-\; H\mathcal{L}(x^*, \lambda^*))(x^k \;-\; x^*)\| \;<\; \gamma \; \|x^k \;-\; x^*\| \; , \; k \in \mathbb{N} \; , \quad (4.23)$$

then the sequences $\{x^k\}_{\mathbb{N}}$ and $\{(x^k, \lambda^k)\}_{\mathbb{N}}$ are well defined and converge linearly to $x^*$ and $(x^*, \lambda^*)$, respectively. The sequence $\{\lambda^k\}_{\mathbb{N}}$ converges R-linearly to $\lambda^*$.

**Proof.** Under assumptions **(B1)-(B3)** the equations (4.18),(4.19) have a solution $(d_x, d_\lambda)$, if $(x^k, \lambda^k)$ is sufficiently close to $(x^*, \lambda^*)$. Observing $\lambda^{k+1} = \lambda^k + d_\lambda$, the solutions can be written in the form

$$d_x = - B_k^{-1} \nabla \mathcal{L}(x^k, \lambda^{k+1}) , \tag{4.24}$$

$$d_\lambda = \big[\nabla h(x^k)^T B_k^{-1} \nabla h(x^k)\big]^{-1} \big[h(x^k) - \nabla h(x^k)^T B_k^{-1} \nabla \mathcal{L}(x^k, \lambda^k)\big] . \tag{4.25}$$

From (4.25) we deduce

$$\lambda^{k+1} = \big[\nabla h(x^k)^T B_k^{-1} \nabla h(x^k)\big]^{-1} \big[h(x^k) - \nabla h(x^k)^T B_k^{-1} \nabla f(x^k)\big] =: W(x^k, B_k) .$$

Setting $A = B_k^{-1}$ in (4.15), we obtain

$$\lambda^* = W(x^*, B_k) .$$

Consequently, we get

$$\begin{aligned} \lambda^{k+1} - \lambda^* &= W(x^k, B_k) - W(x^*, B_k) = \\ &= \big[\nabla h(x^*)^T B_k^{-1} \nabla h(x^*)\big]^{-1} \nabla h(x^*)^T B_k^{-1} (B_k - H\mathcal{L}(x^*, \lambda^*))(x^k - x^*) + w_k , \end{aligned} \tag{4.26}$$

where due to **(B2)** and **(B3)** for some $\kappa > 0$ independent of $k$:

$$w_k \le \kappa \, \|x^k - x^*\|^2 .$$

Combining equations (4.24) and (4.26) and taking **(A2)** into account, we find

$$\begin{aligned} x^{k+1} - x^* &= x^k - x^* - B_k^{-1} \big[\nabla \mathcal{L}(x^k, \lambda^{k+1}) - \nabla \mathcal{L}(x^*, \lambda^*)\big] = \\ &= B_k^{-1} \big[(B_k - H\mathcal{L}(x^*, \lambda^*))(x^k - x^*) - \nabla h(x^*)(\lambda^{k+1} - \lambda^*)\big] + O(\|x^k - x^*\|^2) = \\ &= B_k^{-1} V_k (B_k - H\mathcal{L}(x^*, \lambda^*)(x^k - x^*) + O(\|x^k - x^*\|^2) , \end{aligned} \tag{4.27}$$

where

$$V_k := I - \nabla h(x^*) \big[\nabla h(x^*)^T B_k^{-1} \nabla h(x^*)\big]^{-1} \nabla h(x^*)^T B_k^{-1} .$$

The projection operator $\mathcal{P}$ satisfies

$$V_k \, \mathcal{P}(x^k) = V_k ,$$

and hence, (4.27) implies

$$\|x^{k+1} - x^*\| \le \|B_k^{-1}\| \|V_k\| \|\mathcal{P}(x^k)(B_k - H\mathcal{L}(x^*, \lambda^*)(x^k - x^*)\| + O(\|x^k - x^*\|^2) \tag{4.28}$$

In view of **(B3)**, the assertions can be proved based on the above analysis by using an induction argument.

•

**Remark 4.16 The role of the projection operator**

The condition (4.23) is almost necessary for linear convergence of the sequence $\{x^k\}_{\mathbb{N}}$: Indeed, if we have linear convergence of $\{x^k\}_{\mathbb{N}}$ to $x^*$, then there exists a $0 < \xi < 1$ such that

$$\|\mathcal{P}(x^k)(B_k - H\mathcal{L}(x^*, \lambda^*)(x^k - x^*)\| \; < \; \xi \; \|x^k - x^*\| \; .$$

We note that (4.23) is satisfied under the stronger conditions

$$\|\mathcal{P}(x^k)(B_k - H\mathcal{L}(x^*, \lambda^*)\| \; \leq \; \gamma \; , \tag{4.29}$$

$$\|B_k - H\mathcal{L}(x^*, \lambda^*)\| \; \leq \; \gamma \; . \tag{4.30}$$

In order that (4.29) and (4.30) hold true, it is not necessary that $\{B_k\}_{\mathbb{N}}$ converges to the true Hessian, but rather that the difference $\|B_k - H\mathcal{L}(x^*, \lambda^*)\|$ is kept under control. This requirement gives rise to the following definition:

**Definition 4.17 Bounded deterioration property**

The sequence $\{B_k\}_{\mathbb{N}}$ of matrix approximations within the SQP approach is said to have the bounded deterioration property, if there exist constants $\alpha_i, 1 \leq i \leq 2$, independent of $k \in \mathbb{N}$ such that

$$\|B_{k+1} - H\mathcal{L}(x^*, \lambda^*)\| \; \leq \; (1 + \alpha_1 \sigma_k) \; \|B_k - H\mathcal{L}(x^*, \lambda^*)\| + \alpha_2 \sigma_k \; , \tag{4.31}$$

where

$$\sigma_k \; := \; \max \; \{\|x^{k+1} - x^*\|, \|x^k - x^*\|, \|\lambda^{k+1} - \lambda^*\|, \|\lambda^k - \lambda^*\|\} \; .$$

**Theorem 4.18 Linear convergence in case of the bounded deterioration property**

Assume that the sequence $\{(x^k, \lambda^k)\}_{\mathbb{N}}$ generated by the SQP algorithm and the sequence $\{B_k\}_{\mathbb{N}}$ of symmetric matrix approximations satisfy **(B1)** and the bounded deterioration property (4.31). If $\|x^0 - x^*\|$ and $\|B_0 - H\mathcal{L}(x^*, \lambda^*)\|$ are sufficiently small and $\lambda^0$ is given by (4.16), then the hypotheses of Theorem 3.15 hold true.

**Proof.** The proof can be accomplished by an induction argument and is left as an exercise.

•

Conditions that yield an improvement of linear convergence still depend on the projection of the difference between the approximation and the true Hessian of the Lagrangian:

**Theorem 4.19 Superlinear convergence of the SQP algorithm**

Let $\{(x^k, \lambda^k)\}_{\mathbb{N}}$ be the sequence generated by the SQP algorithm and suppose that conditions **(B1)-(B3)** are satisfied. Further assume that $x^k \to x^*$ as

$k \to \infty$.
Then, $\{x^k\}_{\mathbb{N}}$ converges superlinearly to $x^*$ if and only if the sequence $\{B_k\}_{\mathbb{N}}$ satisfies

$$\lim_{k\to\infty} \frac{\|\mathcal{P}(x^k)(B_k - H\mathcal{L}(x^*,\lambda^*))(x^{k+1}-x^*)\|}{\|x^{k+1}-x^k\|} \; = \; 0 \; . \tag{4.32}$$

Moreover, if (4.32) holds true, then the sequences $\{\lambda^k\}_{\mathbb{N}}$ and $\{(x^k,\lambda^k)\}_{\mathbb{N}}$ converge R-superlinearly and superlinearly, respectively.

**Proof.** The proof is left as an exercise.

### 4.3 Approximations of the Hessian

Approximations $B_k$ of the Hessian of the Lagrangian can be obtained based on the relationship

$$\nabla\mathcal{L}(x^{k+1},\lambda^{k+1}) \; - \; \nabla\mathcal{L}(x^k,\lambda^{k+1}) \; \sim \; H\mathcal{L}(x^{k+1},\lambda^{k+1})(x^{k+1}-x^k) \; . \tag{4.33}$$

Indeed, (4.33) leads to the so-called secant condition

$$B_{k+1}(x^{k+1}-x^k) \; = \; \nabla\mathcal{L}(x^{k+1},\lambda^{k+1}) \; - \; \nabla\mathcal{L}(x^k,\lambda^{k+1}) \; . \tag{4.34}$$

A common strategy is to compute $(x^{k+1},\lambda^{k+1})$ for a given $B_k$ and then to update $B_k$ by means of a rank-one or rank-two update

$$B_{k+1} \; = \; B_k \; + \; U_k \tag{4.35}$$

so that the sequence $\{B_k\}_{\mathbb{N}}$ has the bounded deterioration property.

#### 4.3.1 Rank-two Powell-Symmetric-Broyden update

An update satisfying the bounded deterioration property is the rank-two Powell-Symmetric-Broyden (PSB) formula

$$B_{k+1} \; = \; B_k + \frac{1}{s^T s}\left[(y-B_k s)s^T + s(y-B_k s)^T\right] - \frac{(y-B_k s)^T s}{(s^T s)^2} s s^T \; , \tag{4.36}$$

where

$$s \; = \; x^{k+1} \; - \; x^k \; , \tag{4.37}$$
$$y \; = \; \nabla\mathcal{L}(x^{k+1},\lambda^{k+1}) \; - \; \nabla\mathcal{L}(x^k,\lambda^{k+1}) \; . \tag{4.38}$$

**Theorem 4.20 Properties of the Powell-Symmetric-Broyden update**
Assume that the sequence $\{(x^k,\lambda^k)\}_{\mathbb{N}}$ is generated by the SQP algorithm where the sequence $\{B_k\}_{\mathbb{N}}$ of matrix approximations is obtained by the PSB update formulas (4.36)-(4.38). Assume further that $\|x^0 - x^*\|$ and $\|B_0 - H\mathcal{L}(x^*,\lambda^*)\|$ are sufficiently small and $\lambda^0$ is given by means of (4.16).

Then, the sequence $\{B_k\}_{\mathbb{N}}$ has the bounded deterioration property and the sequence of iterates $\{(x^k, \lambda^k)\}_{\mathbb{N}}$ converges superlinearly to $(x^*, \lambda^*)$ .

**Remark 4.21 Comments on the PSB update**

A drawback of the PSB update is that the matrices $B_k$ are not required to be positive definite. As a consequence, the solvability of the equality constrained QP subproblems is not guaranteed.

### 4.3.2 Rank-two Broyden-Fletcher-Goldfarb-Shanno update

A more suitable two-rank update with regard to postive definiteness is the Broyden-Fletcher-Goldfarb-Shanno (BFGS) update given by

$$B_{k+1} \;=\; B_k + \frac{yy^t}{y^T s} \;-\; \frac{B_k s s^T B_k}{s^T B_k s} \;, \tag{4.39}$$

where $s$ and $y$ are given as in (4.37),(4.38).
In particular, if $B_k$ is positive definite and there holds

$$y^T s \;>\; 0 \;, \tag{4.40}$$

then $B_{k+1}$ is positive definite as well.

**Theorem 4.22 Properties of the BFGS update**

Assume that the sequence $\{(x^k, \lambda^k)\}_{\mathbb{N}}$ is generated by the SQP algorithm where the sequence $\{B_k\}_{\mathbb{N}}$ of matrix approximations is obtained by the PSB update formulas (4.39). Suppose further that $H\mathcal{L}(x^*, \lambda^*)$ and $B_0$ are positive definite. Then, if $\|x^0 - x^*\|$ and $\|B_0 - H\mathcal{L}(x^*, \lambda^*)\|$ are sufficiently small and $\lambda^0$ is given by means of (4.16), the sequence $\{B_k\}_{\mathbb{N}}$ has the bounded deterioration property, (4.32) is satisfied, and the sequence of iterates $\{(x^k, \lambda^k)\}_{\mathbb{N}}$ converges superlinearly to $(x^*, \lambda^*)$ .

**Remark 4.23 Comments on the BFGS update**

The drawback of the BFGS update is that the condition (4.40) is not always satisfied, i.e., there is no guarantee that $B_{k+1}$ turns out to be positive definite.

**Remark 4.24 The Powell-SQP update**

Replacing $y$ in the BFGS-update formula (4.39) by

$$\hat{y} \;=\; \theta y \;+\; (1-\theta) B_k s \quad , \quad \theta \in (0,1] \;, \tag{4.41}$$

is called the Powell-SQP update. In this case, condition (4.40) can always be satisfied, whereas the updates do not longer satisfy the secant condition.

### 4.3.3 The SALSA-SQP update

The SALSA-SQP update relies on an augmented Lagrangian approach, where the objective functional in the NLP is replaced by

$$f_A(x) \;=\; f(x) \;+\; \frac{\eta}{2}\, \|h(x)\|^2 \quad , \quad \eta > 0 \;. \tag{4.42}$$

The associated Lagrangian is usually referred to as the augmented Lagrangian

$$\mathcal{L}_A(x,\lambda) \;=\; \mathcal{L}(x,\lambda) \;+\; \frac{\eta}{2}\,\|h(x)\|^2 \;, \tag{4.43}$$

which also has $(x^*,\lambda^*)$ as a critical point with the Hessian

$$H\mathcal{L}_A(x^*,\lambda^*) \;=\; H\mathcal{L}(x^*,\lambda^*) \;+\; \eta\,\nabla h(x^*)\nabla h(x^*)^T \;. \tag{4.44}$$

If

$$y_A \;=\; \nabla\mathcal{L}_A(x^{k+1},\lambda^{k+1}) \;-\; \nabla\mathcal{L}_A(x^k,\lambda^{k+1}) \;,$$

then

$$y_A \;=\; y \;+\; \eta\nabla h(x^{k+1})h(x^{k+1})^T \;,$$

where $y$ is given by (4.38). Choosing $\eta$ such that

$$y_A^T s \;>\; 0 \;,$$

then $B_{k+1}$ obtained by the BFGS-update formula (with $y$ replaced by $y_A$) is positive definite.
However, since the bounded deterioration property is not guaranteed, a local convergence result is not available.

### 4.4 Reduced Hessian SQP methods

The basic assumption **(A4)** requires $H\mathcal{L}(x^*,\lambda^*)$ to be positive definite only on a particular subspace. Therefore, the reduced Hessian SQP methods approximate the Hessian of the Lagrangian only with respect to that subspace.
We assume that $x^k$ is an iterate for which $\nabla h(x^k)$ has full rank and that $Z_k$ and $Y_k$ are matrices whose columns span the null space of $\nabla h(x^k)^T$ and the range space of $\nabla h(x^k)$, respectively. We further suppose that the columns of $Z_k$ are orthogonal (note that $Z_k$ and $Y_k$ can be computed by an appropriate QR-factorization of $\nabla h(x^k)$).

**Definition 4.25 Reduced Hessian**

Let $(x^k,\lambda^k)$ be a given iterate such that $\nabla h(x^k)$ has full rank. Then, the matrix

$$Z_k^T H\mathcal{L}(x^k,\lambda^k)Z_k \tag{4.45}$$

is called a reduced Hessian for the Lagrangian at $(x^k,\lambda^k)$.

**Remark 4.26 Positive definiteness of the reduced Hessian**

Condition **(A4)** guarantees that the reduced Hessian is positive definite, provided $(x^k,\lambda^k)$ is sufficiently close to $(x^*,\lambda^*)$.

The construction of the update formula for the approximation of the reduced Hessian proceeds very much like the null space approach for solving the KKT system in case of QP problems:
Decomposing the increment $d_x$ according to

$$d_x \;=\; Z_k p_Z \;+\; Y_k p_Y \;, \tag{4.46}$$

the constraint equation of the associated equality constrained QP subproblem reads as

$$\nabla h(x^k)^T Y_k p_Y \;=\; -\; h(x^k) \;. \tag{4.47}$$

Observing **(A2)**, (4.47) has the solution

$$p_Y \;=\; -\; \left[\nabla h(x^k)^T Y_k\right]^{-1} h(x^k) \;. \tag{4.48}$$

The equality constrained QP subproblem then reduces to

$$\text{minimize } \frac{1}{2}\; p_Z^T Z_k^T B_k Z_k p_Z \;+\; (\nabla f(x^k)^T + p_Y^T B_k) Z_k p_Z \;. \tag{4.49}$$

The idea is to use an update of the reduced matrix directly, instead of updating $B_k$ and then computing $Z_k^T B_k Z_k$. Let us assume that we know such a matrix $R_k$ as well as the iterate $(x^k, \lambda^k)$. Then, we first compute $p_Y$ by means of (4.48) and obtain $p_Z$ according to

$$p_Z \;=\; -\; R_k^{-1} Z_k^T (\nabla f(x^k) + B_k p_Y) \;. \tag{4.50}$$

We further set

$$x^{k+1} \;=\; x^k \;+\; d_x \tag{4.51}$$

determine a new multiplier $\lambda^{k+1}$ by

$$\lambda^{k+1} \;=\; -\; \left[\nabla h(x^k)^T \nabla h(x^k)\right]^{-1} \nabla h(x^k)^T \nabla f(x^k) \;, \tag{4.52}$$

and finally update $R_k$ by the BFGS-type update

$$R_{k+1} \;=\; R_k \;+\; \frac{yy^T}{s^T s} \;-\; \frac{R_k s s^T R_k}{s^T R_k s} \;, \tag{4.53}$$

where

$$s \;=\; Z_k^T (x^{k+1} - x^k) \;=\; Z_k^T Z_k p_Z \;, \tag{4.54}$$

$$y \;=\; Z_k^T \left[\nabla \mathcal{L}(x^k + Z_k p_Z, \lambda^k) - \nabla \mathcal{L}(x^k, \lambda^k)\right] \;. \tag{4.55}$$

**Remark 4.27 Determination of $Z_k$**

The null space basis matrices $Z_k$ have to be chosen such that

$$\|Z_k - Z(x^*)\| \;=\; O(\|x^k - x^*\|) \tag{4.56}$$

in order to achieve superlinear convergence.

### 4.5 Convergence monitoring by merit functions

Within the SQP approach global convergence can be achieved by means of appropriately chosen merit functions $M$. The merit functions are chosen in such a way that the solutions of the NLP are unconstrained minimizers of the merit function $M$. At the (k+1)-st iteration step, having determined the Newton increment $d_x$, a suitable steplength $\alpha$ is computed such that

$$M(x^k + \alpha d_x) \;<\; M(x^k) \;. \tag{4.57}$$

The following assumptions will be made to guarantee a steadily decreasing merit function:

**(C1)** The starting point $x^0$ and all subsequent iterates $x^k, k \in \mathbb{N}$, are located in some compact set $K \subset \mathbb{R}^n$.

**(C2)** The columns of $\nabla h(x)$ are linearly independent for all $x \in K$.

The most popular choices of merit functions are augmented Lagrangians and $\ell_p$-norms, $p \geq 1$, of the residual with respect to the KKT conditions.

#### 4.5.1 Augmented Lagrangian merit functions

We consider augmented Lagrangian merit functions $M_A$ of the form

$$M_A(x;\eta) \;=\; f(x) \;+\; h(x)^T \overline{\lambda}(x) \;+\; \frac{\eta}{2}\, \|h(x)\|^2 \;, \tag{4.58}$$

where $\eta > 0$ is a penalty parameter and the multiplier $\overline{\lambda}(x)$ is chosen as the least squares estimate of the optimal multiplier based on the KKT conditions

$$\overline{\lambda}(x) \;=\; -\; \left[\nabla h(x)^T \nabla h(x)\right]^{-1} \nabla h(x)^T \nabla f(x) \;. \tag{4.59}$$

Obviously, $\overline{\lambda}(x^*) = \lambda^*$.
Moreover, under assumptions **(C1),(C2)** we have that $M_A$ and $\overline{\lambda}$ are differentiable and $M_A$ is bounded from below on $K$ for sufficiently large $\eta$.
In particular, for the gradients of $\overline{\lambda}(x)$ and $M_A(x;\eta)$ we obtain

$$\nabla \overline{\lambda}(x) \;=\; -H \nabla \mathcal{L}(x, \overline{\lambda}(x)) \nabla h(x) \left[\nabla h(x)^T \nabla h(x)\right]^{-1} + R_1(x) \;, \tag{4.60}$$

$$\nabla M_A(x;\eta) \;=\; \nabla f(x) + \nabla h(x) \overline{\lambda}(x) + \nabla \overline{\lambda}(x) h(x) + \eta \nabla h(x) h(x) \;, \tag{4.61}$$

where the remainder term $R_1$ in (4.60) is bounded on $K$ and satisfies $R_1(x^*) = 0$, provided $x^*$ satisfies the KKT conditions.

**Theorem 4.28 Properties of augmented Lagrangian merit functions**
Assume that the conditions **(B1),(B2),** and **(C1),(C2)** hold true. Then, for sufficiently large $\eta > 0$ we have:

(i) $x^* \in K$ is a strict local minimum of the NLP if and only if $x^*$ is a strict local minimum of $M_A$.

(ii) If $x$ is not a critical point of the NLP, then $d_x$ is a descent direction for the merit function $M_A$.

**Proof.** Assume that $x^* \in K$ is a feasible point and satisfies the KKT conditions. In view of (4.61) we get

$$\nabla M_A(x^*;\eta) \;=\; 0\;. \tag{4.62}$$

Conversely, if $x^* \in K$ satisfies (4.62), and $\eta$ is chosen sufficiently large, then it follows from **(C1),(C2)** that $h(x^*) = 0$, i.e., $x^*$ is feasible, and $x^*$ satisfies **(A1)**. In order to prove (i) we have to elaborate on the relationship between $HM_A(x^*;\eta)$ and $H\mathcal{L}(x^*,\lambda^*)$.
Reminding that $G(x) \in \mathbb{R}^{n\times m\cdot q_x}$ is the matrix given by

$$G(x) \;:=\; \big(\nabla h_1(x), \nabla h_2(x), ..., \nabla h_m(x), \nabla g_{i_1}(x), ..., \nabla g_{i_{q_x}}(x)\big)\;,$$

we refer to

$$\mathcal{P}(x) \;=\; I \;-\; G(x)\big[G(x)^T G(x)\big]^{-1} G(x)^T \tag{4.63}$$

as the projection onto the null space of $G(x)^T$ and to

$$\mathcal{Q}(x) \;=\; I \;-\; \mathcal{P}(x) \tag{4.64}$$

as the projection onto the range space of $G(x)$.
Then, it follows from (4.61) that

$$\begin{aligned} HM_A(x^*;\eta) \;&=\; H\mathcal{L}(x^*,\lambda^*) - \mathcal{Q}(x^*)H\mathcal{L}(x^*,\lambda^*) - \\ &-\; H\mathcal{L}(x^*,\lambda^*)Q(x^*) \;+\; \eta\nabla h(x^*)\nabla h(x^*)^T\;. \end{aligned} \tag{4.65}$$

Choosing now $y \in \mathbb{R}^n$ and setting $y = \mathcal{Q}(x^*)y + \mathcal{P}(x^*)y$, from (4.65) we get

$$\begin{aligned} y^T HM_A(x^*;\eta)y \;&=\; (\mathcal{P}(x^*)y)^T H\mathcal{L}(x^*,\lambda^*)\mathcal{P}(x^*)y \;- \\ -\;(\mathcal{Q}(x^*)y)^T H\mathcal{L}(x^*,\lambda^*)\mathcal{Q}(x^*)y \;&+\; \eta(\mathcal{Q}(x^*)y)^T\big[\nabla h(x^*)\nabla h(x^*)^T\big]\mathcal{Q}(x^*)y\;. \end{aligned} \tag{4.66}$$

Let us first assume that $x^*$ is a strict local minimum of the NLP. Since $\mathcal{Q}(x^*)$ is in the range of $\nabla h(x^*)$, it follows from **(A2)** that there exists a constant $\mu > 0$ such that

$$(\mathcal{Q}(x^*)y)^T\big[\nabla h(x^*)\nabla h(x^*)^T\big]\mathcal{Q}(x^*)y \;\geq\; \mu\,\|\mathcal{Q}(x^*)y\|^2\;. \tag{4.67}$$

We denote by $\sigma_{min}$ and $\sigma_{max}$ the extreme eigenvalues in the spectrum of $\Sigma(H\mathcal{L}(x^*,\lambda^*))$

$$\sigma_{min} \;:=\; \min\{\sigma \in \Sigma(H\mathcal{L}(x^*,\lambda^*)\} \quad,\quad \sigma_{max} \;:=\; \max\{\sigma \in \Sigma(H\mathcal{L}(x^*,\lambda^*)\}\;,$$

which are both positive in view of **(A4)**. We further define

$$\varepsilon \ := \ \frac{\|\mathcal{Q}(x^*)y\|}{\|y\|}$$

so that

$$\frac{\|\mathcal{P}(x^*)y\|^2}{\|y\|^2} \ = \ 1 \ - \ \varepsilon^2 \ .$$

Dividing (4.66) by $\|y\|^2$ then leads to

$$\frac{y^T HM_A(x^*;\eta)y}{\|y\|^2} \ \geq \ \sigma_{min} \ + \ (\eta\mu - \sigma_{max} - \sigma_{min}) \ \varepsilon^2 \ . \tag{4.68}$$

For sufficiently large $\eta$ the right-hand side in (4.68) is positive, which proves that $x^*$ is a strict local minimum of $M_A$.
Conversely, if $x^*$ is a strict local minimum of $M_A$, then (4.66) is positive for all $y$, i.e., $H\mathcal{L}(x^*,\lambda^*)$ is positive definite on the null space of $\nabla h(x^*)^T$, and hence, $x^*$ is a strict local minimum of the NLP.