Lecture notes for the course "Nonlinear Optimization"

Master Mathematics
Winter term 2025/2026

Univ.-Prof. Dr. Radu Ioan Boţ

Contents

Chapter I

Optimality conditions for differentiable optimization problems

1 Introductory notions

1.1 The Bouligand tangent cone

Let $X \subseteq \mathbb{R}^n$ be a nonempty set, $f : \mathbb{R}^n \to \mathbb{R}$ a given function and the optimization problem

$$\min_{x \in X} f(x). \tag{1.1}$$

An element $x^* \in X$ is called a local minimum of (1.1) if there exists an open neighborhood $B(x^*; \varepsilon) := \{x \in \mathbb{R}^n : \|x - x^*\| < \varepsilon\}$ of x^* such that

$$f(x^*) \leq f(x) \text{ for all } B(x^*; \varepsilon) \cap X. \tag{1.2}$$

Here and in the following, $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^n$. If $f(x^*) \leq f(x)$ for all $x \in X$, then x^* is called a global minimum of (1.1).

In order to characterize the local minima of the problem (1.1), we introduce the so-called Bouligand tangent cone to X at a point $x_0 \in X$.

Definition 1.1 (Bouligand tangent cone) We define the Bouligand tangent cone to the set X at a point $x_0 \in X$ by

$$T_X(x_0) := \left\{ d \in \mathbb{R}^n \,\middle|\, \exists\, (x^k)_{k\geq 0} \subseteq X \;\exists\, (t_k)_{k\geq 0} \downarrow 0, \text{ such that } \frac{x^k - x_0}{t_k} \to d \text{ as } k \to +\infty \right\}.$$

Remark 1.2 (a) Let $X \subseteq \mathbb{R}^n$ and $x_0 \in X$. The Bouligand tangent cone is not empty (since $0 \in T_X(x_0)$), a cone (since for every $d \in T_X(x_0)$ and every $\lambda > 0$ it holds $\lambda d \in T_X(x_0)$), and it is closed. If X is a convex set, then $T_X(x_0)$ is also convex.

(b) If $x_0 \in \operatorname{int} X$, then $T_X(x_0) = \mathbb{R}^n$. Indeed, let $d \in \mathbb{R}^n$. Since $x_0 \in \operatorname{int} X$, there exists $k_0 \geq 1$ such that $x^k := x_0 + \frac{1}{k} d \in X$ for every $k \geq k_0$. Consequently, for $(t_k)_{k\geq 1} := (\frac{1}{k})_{k\geq 1}$, it holds $\frac{x^k - x_0}{t_k} = d$ for every $k \geq k_0$. Thus $d \in T_X(x_0)$.

Proposition 1.3 *Let x^* be a local minimum of (1.1) and $f : \mathbb{R}^n \to \mathbb{R}$ a continuously differentiable function on an open neighbourhood of x^*. Then it holds*

$$\nabla f(x^*)^T d \geq 0 \quad \forall d \in T_X(x^*). \tag{1.3}$$

Proof. Let $d \in T_X(x^*)$. Then there exist sequences $(x^k)_{k\geq 0} \subseteq X$ and $(t_k)_{k\geq 0} \downarrow 0$ such that

$$\frac{x^k - x^*}{t_k} \to d \text{ as } k \to +\infty.$$

Thus $x^k - x^* = \frac{x^k - x^*}{t_k} t_k \to 0$ as $k \to +\infty$. Let $B(x^*; \varepsilon)$ be a neighborhood of x^* on which f is continuously differentiable and for which $f(x) \geq f(x^*)$ for every $x \in B(x^*; \varepsilon) \cap X$. Then there exists a $k_\varepsilon \geq 0$ such that $x^k \in B(x^*; \varepsilon)$ for every $k \geq k_\varepsilon$.

By the Mean Value Theorem, for every $k \geq k_\varepsilon$ there exists $\xi^k \in (x^*, x^k) := \{\lambda x^* + (1-\lambda)x^k : \lambda \in (0,1)\}$ such that

$$f(x^k) - f(x^*) = \nabla f(\xi^k)^T (x^k - x^*).$$

Let $\xi^k := \lambda_k x^* + (1-\lambda_k)x^k$ for $\lambda_k \in (0,1)$. Then, for every $k \geq k_\varepsilon$ we have

$$\left\|\xi^k - x^*\right\| = \left\|\lambda_k x^* + (1-\lambda_k)x^k - x^*\right\| = (1-\lambda_k)\left\|x^k - x^*\right\| \leq \left\|x^k - x^*\right\| \to 0 \text{ as } k \to +\infty,$$

thus $\xi^k \to x^*$ as $k \to +\infty$.

Since the gradient of f is continuous by assumption, we have $\nabla f(\xi^k) \to \nabla f(x^*)$ as $k \to +\infty$ and thus also

$$\nabla f(x^*)^T d = \lim_{k\to\infty} \frac{\nabla f(\xi^k)^T (x^k - x^*)}{t_k} = \lim_{k\to\infty} \frac{f(x^k) - f(x^*)}{t_k} \geq 0,$$

where the last inequality holds since $\frac{f(x^k)-f(x^*)}{t_k} \geq 0$ for every $k \geq k_\varepsilon$. ■

Remark 1.4 We can interpret Proposition 1.3 as follows: at a local minimum x^* of (1.1), there is no tangent direction d to X at x^* for which

$$0 > \nabla f(x^*)^T d = \lim_{t\downarrow 0} \frac{f(x^* + td) - f(x^*)}{t}$$

holds, in other words, for which there exists $t_d > 0$ such that $f(x^* + td) < f(x^*)$ for every $t \in (0, t_d)$.

Definition 1.5 (dual cone) For a cone $K \subseteq \mathbb{R}^n$ we denote its dual cone by

$$K^* := \left\{s \in \mathbb{R}^n : s^T d \geq 0 \, \forall d \in K\right\}.$$

Remark 1.6 The dual cone K^* is the set of vectors that form acute angles with all vectors of the cone K. Proposition 1.3 says that for a local minimum x^* of (1.1) it holds $\nabla f(x^*) \in (T_X(x^*))^*$.

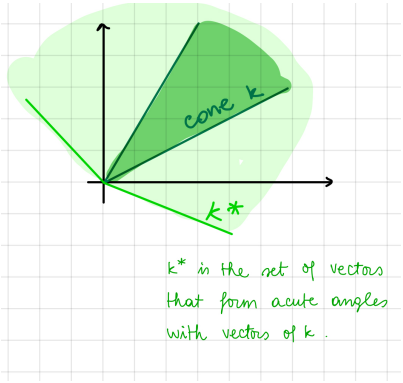

Figure 1.1: Sketch of a dual cone.

Remark 1.7 If X is a convex set and $x_0 \in X$, then

$$T_X(x_0) = \mathrm{cl}(\mathrm{cone}(X - x_0)) = \mathrm{cl}\left(\cup_{\lambda>0}\lambda(X - x_0)\right)$$

and, therefore,

$$-\left(T_X(x_0)\right)^* = N_X(x_0) := \left\{s \in \mathbb{R}^n \middle| \, s^T(x - x_0) \leq 0 \quad \forall x \in X\right\},$$

the so-called normal cone to X at x_0.

If X is convex, then (1.3) is equivalent to $0 \in \nabla f(x^*) + N_X(x^*)$ and further to

$$\nabla f(x^*)^T(x - x^*) \geq 0 \text{ for every } x \in X.$$

For the case $X = \mathbb{R}^n$ we have the following first order optimality condition.

Theorem 1.8 (necessary first order optimality condition) *If x^* is a local minimum of a function $f : \mathbb{R}^n \to \mathbb{R}$ which is differentiable on an open neighbourhood of x^*, then x^* is a critical point of f, in other words,*

$$\nabla f(x^*) = 0. \tag{1.4}$$

Proof. Let $\varepsilon > 0$ such that f is differentiable on $B(x^*; \varepsilon)$ and $f(x) \geq f(x^*)$ for every $x \in B(x^*; \varepsilon)$. Let $t_0 > 0$ such that $x^* - t\nabla f(x^*) \in B(x^*; \varepsilon)$ for all $t \in [0, t_0)$. Then it holds

$$-\|\nabla f(x^*)\|^2 = \nabla f(x^*)^T(-\nabla f(x^*)) = \lim_{t \downarrow 0, 0<t<t_0} \frac{f(x^* - t\nabla f(x^*)) - f(x^*)}{t} \geq 0,$$

therefore, $\nabla f(x^*) = 0$. ■

In the following we consider the general nonlinear optimization problem with inequality and equality constraints

$$\begin{aligned} &\min f(x) \\ &\text{such that } g_i(x) \leq 0, i = 1, \ldots, m \\ &\qquad\quad h_j(x) = 0, i = 1, \ldots, p \\ &\qquad\quad x \in \mathbb{R}^n \end{aligned} \tag{1.5}$$

where $f, g_i, h_j : \mathbb{R}^n \to \mathbb{R}, i = 1, \ldots, m, j = 1, \ldots, p$ are continuously differentiable functions. The set

$$X := \{x \in \mathbb{R}^n \mid g_i(x) \leq 0,\ i = 1 \ldots, m, h_j(x) = 0,\ j = 1, \ldots, p\} \tag{1.6}$$

is called the feasible set of (1.5).

1.2 The linearized tangent cone

We introduce the linearized tangent cone as a "replacement" for the Bouligand tangent cone. The reason for this is that the linearized tangent cone is easy to determine, whereas the Bouligand tangent is generally not.

Definition 1.9 (linearized tangent cone) Let $x_0 \in X$.

a) The constraint $g_i(x) \leq 0$ is said to be active at x_0 if $g_i(x_0) = 0$. We define

$$\mathcal{A}(x_0) := \{i = 1, \ldots, m \mid g_i(x_0) = 0\}$$

as the set of active indices at x_0. We also define the set of inactive indices at x_0 as

$$I(x_0) := \{1, \ldots, m\} \setminus \mathcal{A}(x_0).$$

b) The set

$$T_{\text{lin}}(x_0) := \left\{ d \in \mathbb{R}^n \;\middle|\; \begin{array}{l} \nabla g_i(x_0)^T d \leq 0 \; \forall i \in \mathcal{A}(x_0) \\ \nabla h_j(x_0)^T d = 0 \; \forall j = 1, \ldots, p \end{array} \right\}$$

is called the linearized tangent cone of X at x_0.

The linearized tangent cone is indeed a cone. As we will see in the next sections, in many cases $T_{\text{lin}}(x_0) = T_X(x_0)$.

Remark 1.10 Let $x_0 \in X$. It holds that $T_{\text{lin}}(x_0) = T_{X_{\text{lin}}}(x_0)$, where

$$X_{\text{lin}} := \left\{ x \in \mathbb{R}^n \;\middle|\; \begin{array}{l} g_i(x_0) + \nabla g_i(x_0)^T (x - x_0) \leq 0, i = 1, \ldots, m \\ h_j(x_0) + \nabla h_i(x_0)^T (x - x_0) = 0, j = 1, \ldots, p \end{array} \right\}.$$

Note that $x_0 \in X_{\text{lin}}$.

Lemma 1.11 *Let $x_0 \in X$. Then it holds $T_X(x_0) \subseteq T_{\text{lin}}(x_0)$.*

Proof. Let $x_0 \in X$ and $d \in T_X(x_0)$. Then there exist sequences $(x^k)_{k \geq 0} \subseteq X, (t_k)_{k \geq 0} \downarrow 0$ such that

$$\frac{x^k - x_0}{t_k} \to d \text{ as } k \to \infty.$$

We will first prove that for all $i \in \mathcal{A}(x_0)$, we have $\nabla g_i(x_0)^T d \leq 0$. Indeed, let $i \in \mathcal{A}(x_0)$. By the Mean Value Theorem, for all $k \geq 0$ there exists $\xi^k \in (x_0, x^k)$ which fulfills

$$\nabla g_i(\xi^k)^T (x^k - x_0) = g_i(x^k) - g_i(x_0) = g_i(x^k) \leq 0,$$

since $g_i(x_0) = 0$. Dividing by t_k yields for all $k \geq 0$

$$\nabla g_i(\xi^k)^T \left(\frac{x^k - x_0}{t_k} \right) \leq 0.$$

We let $k \to +\infty$ and, using that ∇g_i is continuous, we obtain

$$\nabla g_i(x_0)^T d \leq 0.$$

Next, we will prove using the same argument that for all $j = 1, \ldots, p$, we have $\nabla h_j(x_0)^T d = 0$. Let $j \in \{1, \ldots, p\}$. By the Mean Value Theorem, for all $k \geq 0$ there exists $\mu^k \in (x_0, x^k)$ such that

$$\nabla h_j(\mu^k)^T (x^k - x_0) = h_j(x^k) - h_j(x_0) = 0,$$

since $h_j(x^k) = h_j(x_0) = 0$. Again, dividing by t_k, we get for all $k \geq 0$

$$\nabla h_j(\mu^k)^T \left(\frac{x^k - x_0}{t_k} \right) = 0.$$

We let $k \to +\infty$ and, using that ∇h_j is continuous, we obtain

$$\nabla h_j(x_0)^T d = 0.$$

This shows that $d \in T_{\text{lin}}(x_0)$. ■

Example 1.12 The equality $T_X(x_0) = T_{\text{lin}}(x_0)$ does not hold in general. Consider the optimization problem

$$\begin{aligned} \min\ & x_1 + x_2^2 \\ \text{such that}\ & g_1(x_1, x_2) := -x_2 \leq 0 \\ & g_2(x_1, x_2) := x_2 - x_1^3 \leq 0 \\ & x = (x_1, x_2) \in \mathbb{R}^2 \end{aligned}$$

Obviously, $x^* := (0, 0)$ is its unique global minimum. We have $\mathcal{A}(x^*) = \{1, 2\}$ and

$$\begin{aligned} \nabla g_1(x_1, x_2) &= (0, -1)^T & \nabla g_1(0, 0) &= (0, -1)^T \\ \nabla g_2(x_1, x_2) &= (-3x_1^2, 1)^T & \nabla g_2(0, 0) &= (0, 1)^T, \end{aligned}$$

so we get $T_{\text{lin}}(x^*) = \mathbb{R} \times \{0\}$.

We easily see that $T_X(x^*) = \mathbb{R}^+ \times \{0\}$. This shows that $T_X(x^*) \subsetneq T_{\text{lin}}(x^*)$. In addition, since $\nabla f(x^*) = (1, 0)^T$, it holds

$$\nabla f(x^*)^T d = d_1 \geq 0 \quad \forall d \in T_X(x^*),$$

however it does not hold that $\nabla f(x^*)^T d = d_1 \geq 0$ for all $d \in T_{\text{lin}}(x^*)$. In conclusion, we cannot replace $T_X(x^*)$ by $T_{\text{lin}}(x^*)$ in (1.3)!

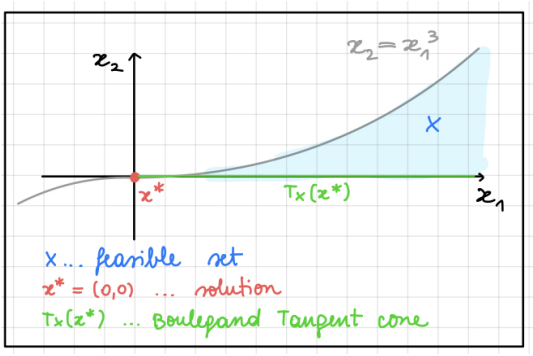

Figure 1.2: Visualization of Example 1.12.

In the following, we will characterize the dual cone of the linearized tangent cone. The following fundamental result will be used for this purpose.

Lemma 1.13 (Lemma of Farkas) *Let $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$. The following statements are equivalent:*

(i) The linear system $Ax = b$ has a solution $x \geqq 0$.

(ii) For all $d \in \mathbb{R}^m$ with $A^T d \geqq 0$ it holds $b^T d \geq 0$.

Theorem 1.14 *Let X be given as in (1.6) and $x_0 \in X$. It holds*

$$-(T_{\text{lin}}(x_0))^* = N_{\text{lin}}(x_0) := \left\{ \sum_{i=1}^{m} \lambda_i \nabla g_i(x_0) + \sum_{j=1}^{p} \mu_j \nabla h_j(x_0) \;\middle|\; \begin{array}{l} \lambda_i \geq 0, i \in \mathcal{A}(x_0) \\ \lambda_i = 0, i \in I(x_0) \\ \mu_j \in \mathbb{R}, j = 1, ..., p \end{array} \right\}. \tag{1.7}$$

Proof. "$\supseteq$": Let $s \in N_{\text{lin}}(x_0)$ and $d \in T_{\text{lin}}(x_0)$. Then

$$(-s)^T d = -\sum_{i=1}^{m} \lambda_i \nabla g_i(x_0)^T d - \sum_{j=1}^{p} \mu_j \nabla h_j(x_0)^T d.$$

Since $\lambda_i \geq 0$ and $\nabla g_i(x_0)^T d \leq 0$ for all $i = 1, \ldots, m$ and $\nabla h_j(x_0)^T d = 0$ for all $j = 1, \ldots, p$, we have

$$(-s)^T d \geq 0.$$

This implies $-s \in (T_{\text{lin}}(x_0))^*$ and therefore $s \in -(T_{\text{lin}}(x_0))^*$.

"$\subseteq$": Take $s \in -(T_{\text{lin}}(x_0))^*$. Then we have, by definition, that $(-s)^T d \geq 0$ for all $d \in T_{\text{lin}}(x_0)$. Define $A \in \mathbb{R}^{n\times(|\mathcal{A}(x_0)|+2p)}$ as

$$A := \big(-\nabla g_{i_1}(x_0), \ldots, -\nabla g_{i_r}(x_0), \nabla h_1(x_0), \ldots, \nabla h_p(x_0), -\nabla h_1(x_0), \ldots, -\nabla h_p(x_0) \big)$$

for $\mathcal{A}(x_0) = \{i_1, \ldots, i_r\}$. We have

$$-s^T d \geq 0 \;\; \forall d \in T_{\text{lin}}(x_0) \Leftrightarrow -s^T d \geq 0 \;\; \forall d \in \mathbb{R}^n \text{ such that } A^T d \geqq 0.$$

By the Lemma of Farkas, there exists $\eta = (\lambda, \mu_1, \mu_2) \in \mathbb{R}_+^{|\mathcal{A}(x_0)|} \times \mathbb{R}_+^p \times \mathbb{R}_+^p$ such that

$$A\eta = -s \Leftrightarrow \sum_{i=1}^{m} -\lambda_i \nabla g_i(x_0) + \sum_{j=1}^{p} (\mu_1)_j \nabla h_j(x_0) + \sum_{j=1}^{p} -(\mu_2)_j \nabla h_j(x_0) = -s$$

$$\Leftrightarrow s = \sum_{i=1}^{m} \lambda_i \nabla g_i(x_0) + \sum_{j=1}^{p} \mu_j \nabla h_j(x_0)$$

where $\lambda_i \geq 0$ for $i \in \mathcal{A}(x_0), \lambda_i = 0$ for $i \in I(x_0)$, and $\mu_j = (\mu_2)_j - (\mu_1)_j \in \mathbb{R}$ for $j = 1, \ldots, p$. ■

Remark 1.15 Since $T_X(x_0) \subseteq T_{\text{lin}}(x_0)$, it holds

$$(T_{\text{lin}}(x_0))^* \subseteq (T_X(x_0))^*.$$

Indeed,

$$s \in (T_{\text{lin}}(x_0))^* \Rightarrow s^T d \geq 0 \; \forall \; d \in T_{\text{lin}}(x_0) \Rightarrow s^T d \geq 0 \; \forall \; d \in T_X(x_0) \Rightarrow s \in (T_X(x_0))^*.$$

Since $(T_{\text{lin}}(x_0))^*$ has the representation (1.7), one would like to have for a local minimum x^* the much nicer condition $\nabla f(x^*) \in (T_{\text{lin}}(x^*))^*$ instead of $\nabla f(x^*) \in (T_X(x^*))^*$. This is in general not the case, as one can see from Example 1.12.

1.3 Karush–Kuhn–Tucker (KKT) optimality conditions

Definition 1.16 (KKT optimality conditions)

(a) The function $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$, given by

$$L(x, \lambda, \mu) = f(x) + \lambda^T g(x) + \mu^T h(x) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \mu_j h_j(x), \tag{1.8}$$

is called the Lagrange function associated with the optimization problem (1.5).

(b) The conditions

$$\begin{cases} \nabla_x L(x, \lambda, \mu) = 0 & (1.9) \\ \lambda \geqq 0, \; g(x) \leqq 0, \; \lambda^T g(x) = 0 & (1.10) \\ h(x) = 0 & (1.11) \end{cases}$$

are called the Karush–Kuhn–Tucker (KKT) optimality conditions of (1.5). It holds that

$$\nabla_x L(x, \lambda, \mu) = \nabla f(x) + \nabla g(x)^T \lambda + \nabla h(x)^T \mu = \nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x) + \sum_{j=1}^{p} \mu_j \nabla h_j(x).$$

(c) An element $(x^*, \lambda^*, \mu^*) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$ fulfilling (1.9) - (1.11) is called a Karush–Kuhn–Tucker (KKT) point of (1.5). The vectors λ^* and μ^* are called the Lagrange multipliers associated with the restrictions $g(x) \leqq 0$ and $h(x) = 0$, respectively. The statement (1.10) is called complementary condition and is equivalent to

$$\lambda_i \geq 0, \; g_i(x^*) \leq 0, \; \lambda_i g_i(x^*) = 0 \text{ for } i = 1, \ldots, m.$$

2 First order necessary and sufficient optimality conditions

2.1 Optimality conditions under (Abadie CQ)

Definition 2.1 An element $x_0 \in X$, where X is given by (1.6), is said to fulfill the Abadie constraint qualification if

$$\text{(Abadie CQ)} \mid T_X(x_0) = T_{\text{lin}}(x_0).$$

Theorem 2.2 *Assume that a local minimum x^* of (1.5) fulfills (Abadie CQ). Then there exist (not necessarily unique) Lagrange multipliers $\lambda^* \in \mathbb{R}^m$ and $\mu^* \in \mathbb{R}^p$ such that (x^*, λ^*, μ^*) is a KKT point of (1.5).*

Proof. From Proposition 1.3, we know that $\nabla f(x^*) \in (T_X(x^*))^*$. Since x^* fulfills (Abadie CQ), we also know that $(T_X(x^*))^* = (T_{lin}(x^*))^*$. From here we get

$$-\nabla f(x^*) \in -(T_{lin}(x^*))^*.$$

From Theorem 1.14, we know that there exist $\lambda_i^* \geq 0$ for $i \in \mathcal{A}(x^*), \lambda_i^* = 0$ for $i \in I(x^*)$, and $\mu_j^* \in \mathbb{R}$ for $j = 1, \ldots, p$, such that

$$-\nabla f(x^*) = \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) + \sum_{j=1}^{p} \mu_j^* \nabla h_j(x^*)$$

or, equivalently,

$$\nabla_x L(x^*, \lambda^*, \mu^*) = 0,$$

where L is the Lagrange function associated with the optimization problem (1.5). In addition, we have $\lambda_i^* \geq 0$, $g_i(x^*) \leq 0$, $\lambda_i^* g_i(x^*) = 0$ for all $i = 1, \ldots, m$. Furthermore, since $x^* \in X$, it holds $h_j(x^*) = 0$ for all $j = 1, \ldots, p$. These last three observations imply that (x^*, λ^*, μ^*) is a KKT point of (1.5). ■

The verification of (Abadie CQ) requires the calculation of the Bouligand tangent cone and of the linearized tangent cone, which can be a complex task. In the next subsections we will introduce sufficient conditions for (Abadie CQ) that are easier to verify. The next corollary shows that (Abadie CQ) is satisfied for optimization problems with linear inequality and equality constraints but general objective functions.

Corollary 2.3 *Let x^* be a local minimum of the optimization problem (1.5), with $g : \mathbb{R}^n \to \mathbb{R}^m, g(x) = Ax - b$, where $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$, and $h : \mathbb{R}^n \to \mathbb{R}^p, h(x) = Cx - d$, where $C \in \mathbb{R}^{p \times n}, d \in \mathbb{R}^p$, i.e.*

$$\begin{aligned} &\min\ f(x). \\ &\text{such that } Ax \leqq b \\ &\qquad\quad Cx = d \end{aligned}$$

Then there exist (not necessarily uniquely defined) Lagrange multipliers $\lambda^* \in \mathbb{R}^m$ and $\mu^* \in \mathbb{R}^p$ such that (x^*, λ^*, μ^*) is a KKT point of (1.5).

Proof. We only need to prove that, for this problem, (Abadie CQ) is satisfied. To do that, it is sufficient to prove that, in this case, we have $X = X_{\text{lin}}$. This implies, according to Remark 1.10, that $T_X(x^*) = T_{X_{\text{lin}}}(x^*) = T_{\text{lin}}(x^*)$, and the statement follows from Theorem 2.2.

Indeed, in this case we have

$$g_i(x) = a_i^T x - b_i, \text{ for } i = 1, \dots, m$$
$$h_j(x) = c_j^T x - d_j, \text{ for } j = 1, \dots, p,$$

and

$$g_i(x^*) + \nabla g_i(x^*)^T(x - x^*) = a_i^T x^* - b_i + a_i^T(x - x^*) = a_i^T x - b_i = g_i(x), \text{ for } i = 1, \dots, m$$
$$h_j(x^*) + \nabla h_j(x^*)^T(x - x^*) = c_j^T x - d_j + c_j^T(x - x^*) = c_j^T x - d_j = h_j(x), \text{ for } j = 1, \dots, p.$$

■

2.2 Optimality conditions under (MFCQ)

Definition 2.4 An element $x_0 \in X$, where X is given by (1.6), is said to fulfill the Mangasarian-Fromovitz constraint qualification if

$$\text{(MFCQ)} \left| \begin{array}{l} (a) \text{ the vectors } \{\nabla h_j(x_0)\}_{j=1}^p \text{ are linearly independent,} \\ (b) \text{ there exists } d \in \mathbb{R}^n \text{ such that } \nabla g_i(x_0)^T d < 0 \forall i \in \mathcal{A}(x_0), \nabla h_j(x_0)^T d = 0 \ \forall j = 1, \dots, p. \end{array} \right.$$

Lemma 2.5 *Let x_0 be a feasible element of (1.5) such that x_0 fulfills (MFCQ), and $d \in \mathbb{R}^n$ the vector which satisfies condition (b) in Definition 2.4. Then there exist $\varepsilon > 0$ and $x : (-\varepsilon, \varepsilon) \to \mathbb{R}^n$ such that*

(a) x in continuously differentiable on $(-\varepsilon, \varepsilon)$;

(b) $x(t) \in X \ \forall t \in [0, \varepsilon)$;

(c) $x(0) = x_0$;

(d) $\dot{x}(0) = d$.

Proof. Define

$$H : \mathbb{R}^p \times \mathbb{R} \to \mathbb{R}^p, \quad (y, t) \mapsto h(x_0 + td + \nabla h(x_0)^T y),$$

where $h = (h_1, ..., h_p) : \mathbb{R}^n \to \mathbb{R}^p$ is the function defining the equality constraint in (1.6) and

$$\nabla h(x_0) = \begin{pmatrix} \nabla h_1(x_0)^T \\ \vdots \\ \nabla h_p(x_0)^T \end{pmatrix} \in \mathbb{R}^{p \times n}$$

is the Jacobi matrix of h at x_0.

It holds that $H(0,0) = h(x_0) = 0$, since x_0 is feasible,

$$\nabla_y H(y,t) = \nabla h(x_0 + td + \nabla h(x_0)^T y)\nabla h(x_0)^T,$$

so $\nabla_y H(0,0) = \nabla h(x_0)\nabla h(x_0)^T \in \mathbb{R}^{p\times p}$.

We claim that $\nabla_y H(0,0)$ is positive definite. Indeed, for all $s \in \mathbb{R}^p$ we have

$$s^T \nabla_y H(0,0)s = s^T \nabla h(x_0)\nabla h(x_0)^T s = (\nabla h(x_0)^T s)^T(\nabla h(x_0)^T s) = \left\|\nabla h(x_0)^T s\right\|^2 \geq 0.$$

Furthermore,

$$s^T \nabla_y H(0,0)s = 0 \Leftrightarrow \nabla h(x_0)^T s = 0 \Leftrightarrow \sum_{j=1}^{p} s_j \nabla h_j(x_0) = 0,$$

which, by condition (a) in (MFCQ), is equivalent to $s = 0$. Thus $\nabla_y H(0,0)$ is positive definite, and therefore invertible.

According to the Implicit Function Theorem for $H(y,t) = 0$ at $(0,0)$, there exist $\varepsilon_0 > 0$ and a continuously differentiable mapping $y : (-\varepsilon_0, \varepsilon_0) \to \mathbb{R}^p$ such that $y(0) = 0$ and

$$H(y(t),t) = 0 \text{ for all } t \in (-\varepsilon_0, \varepsilon_0).$$

By differentiation, we get

$$\nabla_y H(y(t),t)\dot{y}(t) + \nabla_t H(y(t),t) = 0 \text{ for all } t \in (-\varepsilon_0, \varepsilon_0).$$

This yields

$$\nabla_y H(0,0)\dot{y}(0) + \nabla_t H(0,0) = 0 \;\Leftrightarrow\; \dot{y}(0) = -(\nabla_y H(0,0))^{-1}\nabla_t H(0,0),$$

so

$$\dot{y}(0) = -(\nabla y H(0,0))^{-1}\nabla h(x_0)d = 0,$$

where the last equality holds since $\nabla h(x_0)d = 0$ by condition (b) of (MFCQ).

For all $t \in (-\varepsilon_0, \varepsilon_0)$, let

$$x(t) := x_0 + td + \nabla h(x_0)^T y(t),$$

which is obviously continuously differentiable. Furthermore, it holds that $x(0) = x_0$ and

$$\dot{x}(t) = d + \nabla h(x_0)^T \dot{y}(t), \text{ therefore, } \dot{x}(0) = d + \nabla h(x_0)^T \dot{y}(0) = d.$$

and so the statements (c) and (d) are proven.

Next, we note that for all $t \in (0, \varepsilon_0)$ it holds

$$0 = H(y(t),t) = h(x_0 + td + \nabla h(x_0)^T y(t)) = h(x(t)).$$

Let $i \in \{1, \ldots, m\}$. If $g_i(x_0) < 0$, since $x(t) \to x(0) = x_0$ as $t \downarrow 0$, there exists $0 < \varepsilon_i < \varepsilon_0$ such that $g_i(x(t)) < 0$ for all $t \in (0, \varepsilon_i)$.

If $g_i(x_0) = 0$, let $q_i(t) := g_i(x(t))$. For all $t \in (-\varepsilon_0, \varepsilon_0)$ it holds $q'(t) = \nabla g_i(x(t))^T \dot{x}(t)$, and thus $q_i'(0) = \nabla g_i(x(0))^T \dot{x}(0) = \nabla g_i(x_0)^T d < 0$, where the last inequality follows by condition (b) in (MFCQ). On the other hand, we have

$$0 > q_i'(0) = \lim_{t \downarrow 0} \frac{q_i(t) - q_i(0)}{t} = \lim_{t \downarrow 0} \frac{g_i(x(t)) - g_i(x(0))}{t} = \lim_{t \downarrow 0} \frac{g_i(x(t))}{t}.$$

Therefore, there exists $0 < \varepsilon_i < \varepsilon_0$ such that for all $t \in (0, \varepsilon_i)$ it holds

$$\frac{g_i(x(t))}{t} < 0 \text{ or, equivalently, } g_i(x(t)) < 0.$$

We let $\varepsilon := \min\{\varepsilon_1, \ldots, \varepsilon_m\}$. Then, for all $t \in (0, \varepsilon)$ it holds that

$$\begin{aligned} h_j(x(t)) &= 0 \quad \forall j = 1, \ldots, p \\ g_i(x(t)) &\leq 0 \quad \forall i = 1, \ldots, m, \end{aligned}$$

thus $x(t) \in X$. ■

Theorem 2.6 *Assume that a local minimum x^* of (1.5) fulfills (MFCQ). Then there exist (not necessarily unique) Lagrange multipliers $\lambda^* \in \mathbb{R}^m$ and $\mu^* \in \mathbb{R}^p$ such that (x^*, λ^*, μ^*) is a KKT point of (1.5).*

Proof. We show that if $x^* \in X$ satisfies (MFCQ), then it also satisfies (Abadie CQ), and the conclusion will follow from Theorem 2.2. To this end, it is sufficient to prove that $T_{\text{lin}}(x^*) \subseteq T_X(x^*)$, as the opposite inclusion is always true.

Let $d \in T_{\text{lin}}(x^*)$ and $\bar{d} \in \mathbb{R}^n$ be the vector that satisfies condition (b) in (MFCQ). It holds

$$\forall\, i \in \mathcal{A}(x^*) :\ \nabla g_i(x^*)^T d \leq 0 \text{ and } \nabla g_i(x^*)^T \bar{d} < 0,$$

and

$$\forall\, j = 1 \ldots p :\ \nabla h_j(x^*)^T d = 0 \text{ and } \nabla h_j(x^*)^T \bar{d} = 0.$$

For all $\tau > 0$ we define $d(\tau) := d + \tau \bar{d}$ and see that

$$\forall\, i \in \mathcal{A}(x^*) :\ \nabla g_i(x^*)^T d(\tau) < 0 \text{ and } \forall\, j = 1 \ldots p :\ \nabla h_j(x^*)^T d(\tau) = 0.$$

This implies that $d(\tau)$ fulfills (MFCQ) for all $\tau > 0$.

We fix $\tau > 0$. By Lemma 2.5, there exist $\varepsilon > 0$ and a continuously differentiable mapping $x : (-\varepsilon, \varepsilon) \to \mathbb{R}^n$ such that

$$x(t) \in X \ \forall t \in [0, \epsilon), \quad x(0) = x^*, \quad \text{and} \quad \dot{x}(0) = d(\tau).$$

For $k \geq 2$, we denote $t_k := \frac{\varepsilon}{k}$ and $x_k := x(t_k) \in X$. Then

$$\lim_{k \to +\infty} \frac{x_k - x^*}{t_k} = \lim_{k \to +\infty} \frac{x(t_k) - x(0)}{t_k} = \dot{x}(0) = d(\tau),$$

which proves that $d(\tau) \in T_X(x^*)$. Since $d(\tau) \to d$ as $\tau \downarrow 0$ and $T_X(x^*)$ is closed, it yields $d \in T_X(x^*)$. ■

2.3 Optimality conditions under (LICQ)

Definition 2.7 An element $x_0 \in X$, where X is given by (1.6), is said to fulfill the Linear Independence constraint qualification if

$$\text{(LICQ)} \mid \text{the vectors } \{\nabla g_i(x_0)\}_{i\in\mathcal{A}(x_0)} \cup \{\nabla h_j(x_0)\}_{j=1}^p \text{ are linearly independent.}$$

Theorem 2.8 *Assume that a local minimum x^* of (1.5) fulfills (LICQ). Then there exist uniquely defined Lagrange multipliers $\lambda^* \in \mathbb{R}^m$ and $\mu^* \in \mathbb{R}^p$ such that (x^*, λ^*, μ^*) is a KKT point of (1.6).*

Proof. First, we will show that x^* fulfils (MFCQ).

Since

$$\{\nabla h_j(x^*)\}_{j=1}^p \subseteq \{\nabla g_i(x^*)\}_{i\in\mathcal{A}(x^*)} \cup \{\nabla h_j(x^*)\}_{j=1}^p,$$

the vectors $\nabla h_j(x^*), j = 1,\dots,p$, are linearly independent, thus condition (a) in (MFCQ) is satisfied.

Note that $|\mathcal{A}(x^*)| + p \le n$. Let $A \in \mathbb{R}^{n\times n}$ be a regular matrix defined as

$$A := \begin{pmatrix} \nabla g_i(x^*)^T|_{i\in\mathcal{A}(x)} \\ \nabla h_j(x^*)^T|_{j=1,\dots,p} \\ \text{any rows that complete the set of vectors to a basis of } \mathbb{R}^n \end{pmatrix}.$$

and

$$b := (\underbrace{-1,\dots,-1}_{|\mathcal{A}(x^*)| \text{ many}}, \underbrace{0,\dots,0}_{p \text{ many}}, \underbrace{\text{arbitrary real numbers}}_{n-p-|\mathcal{A}(x^*)| \text{ many}})^T \in \mathbb{R}^n.$$

Since A is regular, there exists a unique $d \in \mathbb{R}^n$ such that $Ad = b$, which implies that

$$\nabla g_i(x^*)^T d = -1 < 0 \ \ \forall i \in \mathcal{A}(x^*) \quad \text{and} \quad \nabla h_j(x^*)^T d = 0 \ \ \forall j = 1,\dots,p.$$

Thus, condition (b) in (MFCQ) is also satisfied, and, according to Theorem 2.6, there exist Lagrange multipliers $\lambda^* \in \mathbb{R}^n_+$ and $\mu^* \in \mathbb{R}^p$ such that (x^*, λ^*, μ^*) is a KKT point of (1.6).

To prove the uniqueness, of the Lagrange multipliers, we note that for all inactive indices $i \in I(x^*)$ we have $g_i(x^*) < 0$, and so $\lambda_i^* = 0$. Thus,

$$\nabla f(x^*) = -\sum_{i\in\mathcal{A}(x^*)} \lambda_i^* \nabla g_i(x^*) - \sum_{j=1}^p \mu_j^* \nabla h_j(x^*),$$

and by (LICQ) it follows that $\lambda_i^*, i \in \mathcal{A}(x^*)$ and $\mu_j, j = 1,\dots,p$ are unique. ∎

So far, out of the constraint qualifications we have introduced, LICQ is the easiest to verify. There are optimization problems for which (LICQ) is not satisfied, but (MFCQ) or (Abadie CQ) are.

2.4 Optimality conditions for convex optimization problems

In this subsection, we study the optimization problem (1.5) under the following additional assumptions:

(a) $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function;

(b) $g_i : \mathbb{R}^n \to \mathbb{R}, i = 1, \dots, m$, are convex functions;

(c) $h : \mathbb{R}^n \to \mathbb{R}^p$ is an affine function, i.e. $h : \mathbb{R}^n \to \mathbb{R}^p, h(x) = Ax - b$, with $A \in \mathbb{R}^{p\times n}$ and $b \in \mathbb{R}^p$.

The optimization problem (1.5) becomes

$$\begin{aligned} &\min\ f(x). \\ &\text{such that } g_i(x) \le 0,\ i = 1, \dots, m \\ &\qquad Ax = b \\ &\qquad x \in \mathbb{R}^n \end{aligned} \tag{2.1}$$

Under this assumptions, the feasible set $X = \{x \in \mathbb{R}^n \mid g_i(x) \le 0, i = 1, ..., m, Ax = b\}$ is convex.

Definition 2.9 We say that Slater's constraint qualification is fulfilled for the convex optimization problem (2.1) if

$$\text{(Slater CQ)} \mid \exists\, x' \in \mathbb{R}^n \text{ such that } g_i(x') < 0,\ i = 1, \dots, m, \text{ and } Ax' = b.$$

Unlike all the constraint qualifications we have considered so far, (Slater CQ) is a global assumption in the sense that it does not depend on a priori given feasible elements.

Theorem 2.10 *Let x^* be a (local = global) minimum of (2.1) and (Slater CQ) be fulfilled. Then there exist (not necessarily unique) Lagrange multipliers $\lambda^* \in \mathbb{R}^m$ and $\mu^* \in \mathbb{R}^p$ such that (x^*, λ^*, μ^*) is a KKT point of (2.1) .*

Proof. We will show that x^* fulfils (Abadie CQ) and the conclusion will follow from Theorem 2.2. In particular, we will show that $T_{\text{lin}}(x^*) \subseteq T_X(x^*)$. Let

$$d \in T_{\text{lin}}(x^*) = \left\{ \tilde{d} \in \mathbb{R}^n \;\middle|\; \nabla g_i(x^*)^T \tilde{d} \le 0, i \in \mathcal{A}(x^*), A\tilde{d} = 0 \right\}.$$

(Slater CQ) guarantees that there exists $x' \in X$ such that $g_i(x') < 0, i = 1, \dots, m$, and $Ax' = b$. Let $d' := x' - x^*$. Due to the gradient inequality for convex functions, we have for all $i \in \mathcal{A}(x^*)$

$$\nabla g_i(x^*)^T d' = \nabla g_i(x^*)^T (x^* + d' - x^*) \le g_i(x^* + d') - g_i(x^*) = g_i(x') - g_i(x^*) = g_i(x') < 0.$$

In addition, it holds

$$Ad' = Ax' - Ax^* = b - b = 0.$$

Define $d(\tau) := d + \tau d'$ for all $\tau > 0$. For all $\tau > 0$ it holds

$$\nabla g_i(x^*)^T d(\tau) = \nabla g_i(x^*)^T d + \tau \nabla g_i(x^*)^T d' < 0 \ \text{ for all } i \in \mathcal{A}(x^*),$$

and

$$Ad(\tau) = Ad + \tau Ad' = 0.$$

Next, we will show that $d(\tau) \in T_X(x^*)$ for all $\tau > 0$. Indeed, define

$$x^k := x^* + \frac{1}{k} d(\tau) \text{ and } t_k := \frac{1}{k} \text{ for all } k \geq 1,$$

which gives

$$\frac{x^k - x^*}{t_k} = d(\tau) \to d(\tau) \text{ as } k \to +\infty.$$

To finish the proof, it remains to show that

$$x^k \in X = \{x \in \mathbb{R}^n \mid g_i(x) \leq 0, i = 1, ..., m, Ax = b\}$$

for all k large enough.

First, we have for all $k \geq 1$

$$Ax^k = Ax^* + \frac{1}{k} Ad(\tau) = b.$$

Next, let $i \in \mathcal{A}(x^*)$. By the Mean Value Theorem, for all $k \geq 1$ there exits $\xi^k \in (x^*, x^k)$ such that

$$g_i(x^k) = g_i(x^k) - g_i(x^*) = \nabla g_i(\xi^k)^T (x^k - x^*) = \frac{1}{k} \nabla g_i(\xi^k)^T d(\tau).$$

Letting k convergence to infinity, we get

$$0 > \nabla g_i(x^*)^T d(\tau) = \lim_{k \to +\infty} \nabla g_i(\xi^k)^T d(\tau) = \lim_{k \to +\infty} k g_i(x^k).$$

This implies that there exists $k_0^i \geq 1$ such that $g_i(x^k) < 0$ for all $k \geq k_0^i$.

Lastly, let $i \in I(x^*)$. Since $g_i(x^*) < 0$ and $x^k \to x^*$ as $k \to \infty$, there exists $k_0^i \geq 1$ such that $g_i(x^k) < 0$ for all $k \geq k_0^i$.

Then, for

$$k_0 := \max_{i=1,...,m} \{k_0^i\},$$

it holds $x^k \in X$ for all $k \geq k_0$, which implies $d(\tau) \in T_X(x^*)$. Since $d = \lim_{\tau \downarrow 0} d(\tau)$ and $T_X(x^*)$ is closed, we obtain $d \in T_X(x^*)$. ■

We close this section with a result that shows that in the convex setting the KKT optimality conditions are also sufficient for optimality.

Theorem 2.11 *Let (x^*, λ^*, μ^*) be a KKT point of (2.1). Then x^* is a (local =global) minimum of (2.1).*

Proof. Let $x \in X$. Using the gradient inequality for convex functions, we have:

$$f(x) \geq f(x^*) + \nabla f(x^*)^T(x - x^*) = f(x^*) + \left(-\sum_{i=1}^{n} \lambda_i^* \nabla g_i(x^*)^T - \sum_{j=1}^{p} \mu_j^* a_j^T \right)(x - x^*)$$
$$= f(x^*) - \sum_{i \in \mathcal{A}(x^*)} \lambda_i^* \nabla g_i(x^*)^T(x - x^*),$$

where a_j^T denotes the j-th row of A. The second equality takes into account that, for i inactive, $\lambda_i = 0$, and $Ax - Ax^* = b - b = 0$. By using again the gradient inequality, we obtain further

$$f(x) \geq f(x^*) - \sum_{i \in \mathcal{A}(x^*)} \lambda_i^*(g_i(x) - g_i(x^*)) = f(x^*) - \sum_{i \in \mathcal{A}(x^*)} \lambda_i^*(g_i(x)) \geq f(x^*).$$

This proves the statement. ■

3 Second order necessary and sufficient optimality conditions

In this section, we will formulate second order necessary and sufficient conditions for both unconstrained and constrained optimization problems.

3.1 The unconstrained case

Theorem 3.1 *Let x^* be a local minimum of $f : \mathbb{R}^n \to \mathbb{R}$ and f be twice continuously differentiable on a open neighborhood $B(x^*; \varepsilon)$ of x^*. Then the Hessian $\nabla^2 f(x^*) \in \mathbb{R}^{n \times n}$ is positive semidefinite.*

Proof. We assume that there exists $d \in \mathbb{R}^n$ such that $d^T \nabla^2 f(x^*) d < 0$. Since the map $x \mapsto d^T \nabla^2 f(x) d$ is continuous at x^*, there exists $\alpha > 0$ such that for all $t \in [0, \alpha]$ it holds

$$x^* + td \in B(x^*; \varepsilon) \text{ and } d^T \nabla^2 f(x^* + td) d < 0.$$

Let $t \in [0, \alpha]$. According to Taylor's Theorem for $s \mapsto f(x^* + sd)$, there exists $\xi_t \in (0, t)$ such that

$$f(x^* + td) = f(x^*) + t \nabla f(x^*)^T d + \frac{1}{2} t^2 d^T \nabla^2 f(x^* + \xi_t d) d.$$

By Theorem 1.8, we have

$$f(x^* + td) = f(x^*) + \frac{1}{2} t^2 d^T \nabla^2 f(x^* + \xi_t d) d < f(x^*),$$

which implies that $f(x^* + td) < f(x^*)$ for all $t \in [0, \alpha]$. This contradicts the local minimality of x^*. ■

Example 3.2 (a) For $f : \mathbb{R} \to \mathbb{R}, f(x) = -x^2$, we have $f'(x^*) = 0$ if and only if $x^* = 0$. Since $f''(x^*) = -2 < 0$, Theorem 3.1 allows us to conclude that f has no local minima.

(b) The conditions $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ is positive semidefinite are not sufficient for x^* to be a local minimum of f. Let $f : \mathbb{R}^2 \to \mathbb{R},\ f(x_1, x_2) = x_1^2 - x_2^3$. It holds

$$\nabla f(x_1, x_2) = \begin{pmatrix} 2x_1 \\ -3x_2^2 \end{pmatrix} \text{ and } \nabla^2 f(x_1, x_2) = \begin{pmatrix} 2 & 0 \\ 0 & -6x_2 \end{pmatrix}.$$

We have

$$\nabla f(x^*) = 0 \ \Leftrightarrow \ x^* = (0, 0)^T,$$

and

$$\nabla^2 f(x^*) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$$

is positive semidefinite. But x^* is not a local minimum of f, since, for all $\varepsilon > 0$, there exists $(0, \varepsilon/2)^T \in B(x^*; \varepsilon)$ with

$$f(0, \varepsilon/2) = -\frac{\varepsilon^3}{8} < f(x^*) = 0.$$

A sufficient optimality condition is given by the following theorem.

Theorem 3.3 *Let f be twice continuous differentiable on a open neighborhood $B(x^*; \varepsilon)$ of an element $x^* \in \mathbb{R}^n$ such that:*

a) $\nabla f(x^*) = 0$, *and*

b) $\nabla^2 f(x^*)$ *is positive definite.*

Then x^* is a strict local minimum of f, i.e., there exists $\delta > 0$ such that

$$f(x^*) < f(x) \quad \forall x \in B(x^*; \delta) \setminus \{x^*\}.$$

Proof. Since $\nabla^2 f(x^*)$ is positive definite, its smallest eigenvalue $\lambda_{min}(\nabla^2 f(x^*))$ is positive and for all $d \in \mathbb{R}^n$ it holds

$$d^T \nabla^2 f(x^*) d \geq \lambda_{min}(\nabla^2 f(x^*)) \, \|d\|^2 .$$

By the continuity of $x \mapsto \nabla^2 f(x)$ at x^*, there exists $0 < \delta < \varepsilon$ such that for all $d \in \mathbb{R}^n$ with $\|d\| < \delta$ it holds

$$\left\| \nabla^2 f(x^* + d) - \nabla^2 f(x^*) \right\| < \frac{\lambda_{min}(\nabla^2 f(x^*))}{2},$$

where the norm on the left-hand side denotes the operator norm of a matrix[1].

[1]The *operator norm* of a matrix is defined as $\|\cdot\| : \mathbb{R}^{n \times n} \to \mathbb{R}, A \mapsto \max_{x \neq 0} \frac{\|Ax\|}{\|x\|}$.

Choose $d \in \mathbb{R}^n$ such that $\|d\| < \delta$. Taylor's Theorem for $s \mapsto f(x^* + sd)$ gurantees that there exists a $\xi_d \in (0,1)$ such that

$$\begin{aligned} f(x^*+d) &= f(x^*) + \nabla f(x^*)^T d + \frac{1}{2} d^T \nabla^2 f(x^* + \xi_d d) d \\ &= f(x^*) + \frac{1}{2} d^T \nabla^2 f(x^* + \xi_d d) d \\ &= f(x^*) + \frac{1}{2} d^T \nabla^2 f(x^*) d + \frac{1}{2} d^T \nabla^2 f(x^* + \xi_d d) d - \frac{1}{2} d^T \nabla^2 f(x^*) d \\ &= f(x^*) + \frac{1}{2} d^T \nabla^2 f(x^*) d + \frac{1}{2} d^T \Big(\nabla^2 f(x^* + \xi_d d) - \nabla^2 f(x^*) \Big) d. \end{aligned}$$

The Cauchy-Schwarz inequality and the fact that $\|\xi_d d\| \leq \|d\| < \delta$ yield

$$\begin{aligned} f(x^*+d) &\geq f(x^*) + \frac{1}{2} \lambda_{min}(\nabla^2 f(x^*)) \|d\|^2 - \frac{1}{2} \|d\|^2 \left\| \nabla^2 f(x^* + \xi_d d) - \nabla^2 f(x^*) \right\| \\ &\geq f(x^*) + \frac{1}{2} \lambda_{min}(\nabla^2 f(x^*)) \|d\|^2 - \frac{1}{4} \lambda_{min}(\nabla^2 f(x^*)) \|d\|^2 \\ &= f(x^*) + \frac{1}{4} \lambda_{min}(\nabla^2 f(x^*)) \|d\|^2 . \end{aligned}$$

Plugging in $d := x - x^*$, we get for all $x \in B(x^*; \delta) \setminus \{x^*\}$

$$f(x) \geq f(x^*) + \frac{1}{4} \lambda_{min}(\nabla^2 f(x^*)) ||x - x^*||^2 > f(x^*).$$

■

Remark 3.4 If x^* is a strict local miminum of f and f is twice continuously differentiable on a open neighbourhood of x^*, then $\nabla f(x^*) = 0$, but the Hessian $\nabla^2 f(x^*)$ is not necessarily positive definite. Indeed, for $f : \mathbb{R} \to \mathbb{R}, f(x) = x^4$, $x^* = 0$ is a strict (local = global) minimum of f, $f'(x^*) = 0$, however, $f''(x^*) = 0$.

3.2 The constrained case

In this subection, we study the constrained optimization problem (1.5) under the assumption that the functions $f, g_i, i = 1, ..., m$, and $h_j, j = 1, ..., p$, are twice continuously differentiable. For a given KKT point (x^*, λ^*, μ^*) of (1.5), we consider the following two subsets of $\mathcal{A}(x^*)$:

$$\begin{aligned} \mathcal{A}_0(x^*) &:= \{i \in \mathcal{A}(x^*) \mid \lambda_i^* = 0\} - \text{ the so-called set of weak active indices;} \\ \mathcal{A}_>(x^*) &:= \{i \in \mathcal{A}(x^*) \mid \lambda_i^* > 0\} - \text{ the so-called the set of strong active indices.} \end{aligned}$$

Further, we introduce the following subset of $T_{\text{lin}}(x^*)$

$$T_2(x^*) := \left\{ d \in \mathbb{R}^n \;\middle|\; \begin{array}{l} \nabla g_i(x^*)^T d = 0 \;\forall i \in \mathcal{A}_>(x^*) \\ \nabla g_i(x^*)^T d \leq 0 \;\forall i \in \mathcal{A}_0(x^*) \\ \nabla h_j(x^*)^T d = 0 \;\forall j = 1, \ldots, p \end{array} \right\}$$

Theorem 3.5 *Let x^* be a local minimum of (1.5) which satisfies (LICQ). Let $\lambda^* \in \mathbb{R}^m$ and $\mu^* \in \mathbb{R}^p$ be the (according to Theorem 2.8) uniquely defined Lagrange multipliers such that (x^*, λ^*, μ^*) is a KKT point of (1.5). Then it holds*

$$d^T \nabla^2_{xx} L(x^*, \lambda^*, \mu^*) d \geq 0 \quad \forall d \in T_2(x^*),$$

in other words, $\nabla^2_{xx} L(x^*, \lambda^*, \mu^*)$ is positive semidefinite on the cone $T_2(x^*)$.

Proof. Let $d \in T_2(x^*), d \neq 0$. We split $\mathcal{A}_0(x^*)$ into the following two sets of indices:

$$\mathcal{A}_0^<(x^*) := \{i \in \mathcal{A}_0(x^*) \mid \nabla g_i(x^*)^T d < 0\}$$
$$\mathcal{A}_0^=(x^*) := \{i \in \mathcal{A}_0(x^*) \mid \nabla g_i(x^*)^T d = 0\}.$$

For all $x \in \mathbb{R}^n$, we define

$$\tilde{g}(x) := \begin{pmatrix} g_i(x)|_{i \in I(x^*)} \\ g_i(x)|_{i \in \mathcal{A}_0^<(x^*)} \end{pmatrix} \text{ and } \tilde{h}(x) := \begin{pmatrix} h_j(x)|_{j=1,\ldots,p} \\ g_i(x)|_{i \in \mathcal{A}_>(x^*)} \\ g_i(x)|_{i \in \mathcal{A}_0^=(x^*)} \end{pmatrix}.$$

Furthermore, we define

$$\widetilde{X} := \{x \in \mathbb{R}^n \mid \tilde{g}(x) \leqq 0, \tilde{h}(x) = 0\} \subseteq X.$$

It is easy to see that $x^* \in \widetilde{X}$. Further, we will demonstrate that x^* satisfies (MFCQ) for $\widetilde{X}$ and the vector d fixed above.

(a) The set

$$\{\nabla h_j(x^*)\}_{j=1}^p \cup \{\nabla g_i(x^*)\}_{i \in \mathcal{A}_>(x^*)} \cup \{\nabla g_i(x^*)\}_{i \in \mathcal{A}_0^=(x^*)}$$

is linearly independent, as it is a subset of

$$\{\nabla h_j(x^*)\}_{j=1}^p \cup \{\nabla g_i(x^*)\}_{i \in \mathcal{A}(x^*)}$$

which is linearly independent by (LICQ).

(b) The vector d fulfills

$$\begin{cases} \nabla g_i(x^*)^T d < 0 & \forall i \in \mathcal{A}_0^<(x^*) \quad \text{(by the definition of } \mathcal{A}_0^<(x^*)) \\ \nabla h_j(x^*)^T d = 0 & \forall j = 1, \ldots, p \quad \text{(since } d \in T_2(x^*)) \\ \nabla g_i(x^*)^T d = 0 & \forall i \in \mathcal{A}_>(x^*) \quad \text{(since } d \in T_2(x^*)) \\ \nabla g_i(x^*)^T d = 0 & \forall i \in \mathcal{A}_0^=(x^*) \quad \text{(by the definition of } \mathcal{A}_0^=(x^*) \end{cases},$$

therefore, by Lemma 2.5, there exist $\varepsilon > 0$ and a twice continuously differentiable map $x : (-\varepsilon, \varepsilon) \to \mathbb{R}^n$ (this follows from the Implicit Function Theorem in the proof of Lemma 2.5 by using that h is twice continuously differentiable) such that $x(0) = x^*$, $\dot{x}(0) = d$ and $x(t) \in \widetilde{X}$ for all $t \in [0, \varepsilon)$.

Define

$$\varphi : (-\varepsilon, \varepsilon) \to \mathbb{R}, \quad \varphi(t) = L(x(t), \lambda^*, \mu^*),$$

where L is the Lagrangian associated with the optimization problem (1.5). Note that φ is twice continuously differentiable. For all $t \in (-\varepsilon, \varepsilon)$, we have

$$\begin{aligned}\dot{\varphi}(t) &= \nabla_x L(x(t), \lambda^*, \mu^*)^T \dot{x}(t), \text{ and}\\ \ddot{\varphi}(t) &= \dot{x}(t)^T \nabla_{xx} L(x(t), \lambda^*, \mu^*)^T \dot{x}(t) + \nabla_x L(x(t), \lambda^*, \mu^*)^T \ddot{x}(t).\end{aligned}$$

Since $\nabla_x L(x^*, \lambda^*, \mu^*) = 0$, this gives

$$\dot{\varphi}(0) = 0 \text{ and } \ddot{\varphi}(0) = d^T \nabla_{xx} L(x^*, \lambda^*, \mu^*) d.$$

Note that for all $t \in [0, \varepsilon)$ it holds

- $\sum_{i \in \mathcal{A}_>(x^*)} \lambda_i^* g_i(x(t)) = 0$, since $x(t) \in \widetilde{X}$ and, therefore, $g_i(x(t)) = 0$ for every $i \in \mathcal{A}_>(x^*)$;
- $\sum_{i \in \mathcal{A}_0(x^*)} \lambda_i^* g_i(x(t)) = 0$, since $\lambda_i^* = 0$ for every $i \in \mathcal{A}_0(x^*)$;
- $\sum_{i \in I(x^*)} \lambda_i^* g_i(x(t)) = 0$, since $\lambda_i^* = 0$ for every $i \in I(x^*)$;
- $\sum_{j=1}^p \mu_i^* h_j(x(t)) = 0$, since $x(t) \in \widetilde{X}$ and, therefore, $h_j(x(t)) = 0$ for every $j = 1, \ldots, p$.

Therefore, by definition of the Lagrangian,we have for all $t \in [0, \varepsilon)$

$$\begin{aligned}\varphi(t) &= L(x(t), \lambda^*, \mu^*)\\ &= f(x(t)) + \sum_{i \in \mathcal{A}_>(x^*)} \lambda_i^* g_i(x(t)) + \sum_{i \in \mathcal{A}_0(x^*)} \lambda_i^* g_i(x(t)) + \sum_{i \in I(x^*)} \lambda_i^* g_i(x(t)) + \sum_{j=1}^p \mu_i^* h_j(x(t))\\ &= f(x(t))\end{aligned}$$

Since $x(0) = x^*$ is a local minimum of f, there exists $\delta > 0$ such that

$$f(z) \geq f(x^*) \quad \text{for all } z \in B(x^*; \delta) \cap X.$$

Since x is continuous at 0, there exists $\alpha \in (0, \varepsilon)$ such that for all $t \in [0, \alpha)$ it holds $x(t) \in B(x^*; \delta)$. This implies that for all $t \in [0, \alpha)$ it holds $x(t) \in B(x^*; \delta) \cap \widetilde{X} \subseteq B(x^*; \delta) \cap X$, thus

$$f(x(t)) \geq f(x^*).$$

In other words,

$$\varphi(t) \geq \varphi(0) \quad \forall t \in [0, \alpha). \tag{3.1}$$

We want to prove that $\ddot{\varphi}(0) \geq 0$. Assume that $\ddot{\varphi}(0) < 0$. Since φ is twice continuously differentiable, there exists $\beta \in (0, \alpha)$ such that $\ddot{\varphi}(t) < 0$ for all $t \in [0, \beta)$. By Taylor's Theorem, there exists $t_\beta \in (0, \beta)$ such that

$$\varphi(\beta) = \varphi(0) + \beta \dot{\varphi}(0) + \frac{1}{2}\beta^2 \ddot{\varphi}(t_\beta) = \varphi(0) + \frac{1}{2}\beta^2 \ddot{\varphi}(t_\beta) < \varphi(0).$$

This contradicts (3.1). ■

The following result provides a second order sufficient optimality condition for the constrained optimization problem (1.5).

Theorem 3.6 *Let (x^*, λ^*, μ^*) be a KKT point of (1.5) such that for all $d \in T_2(x^*) \setminus \{0\}$, it holds*

$$d^T \nabla^2_{xx} L(x^*, \lambda^*, \mu^*) d > 0 \quad \forall d \in T_2(x^*) \setminus \{0\},$$

i.e., $\nabla_{xx} L(x^*, \lambda^*, \mu^*)$ is positive definite on the cone $T_2(x^*)$. Then x^* is a strict local minimum of (1.5), i.e., there exists $\delta > 0$ such that

$$f(x) > f(x^*) \quad \forall x \in B(x^*; \delta) \cap X \setminus \{x^*\}.$$

Proof. Assume that x^* is not a strict local minimum. of (1.5). Then, for all $k \geq 1$

$$\text{there exists } x^k \in B\left(x^*; \frac{1}{k}\right) \cap X \setminus \{x^*\} \text{ such that } f(x^k) \leq f(x^*). \tag{3.2}$$

We define for all $k \geq 1$

$$d^k := \frac{1}{||x^k - x^*||}(x^k - x^*).$$

Then, $\|d_k\| = 1$ for all $k \geq 1$ and, as $(d^k)_{k\geq 1}$ is a bounded sequence, it has subsequence $(d^{k_l})_{l\geq 1}$ that convergences to an element $d^* \in \mathbb{R}^n$ with $\|d^*\| = 1$. We will show that $d^* \in T_2(x^*) \setminus \{0\}$ and that $(d^*)^T \nabla^2_{xx} L(x^*, \lambda^*, \mu^*) d^* \leq 0$, which will lead to a contradiction and thus prove the theorem.

First, we will prove that $d^* \in T_{\text{lin}}(x^*)$.

Let $j \in \{1, \ldots, p\}$. By the Mean Value Theorem, for all $k \geq 1$ we know that there exists $\xi^k \in (x^*, x^k)$ such that, since $x^k, x^* \in X$, the following holds

$$0 = h_j(x^k) - h_j(x^*) = \nabla h_j(\xi^k)^T (x^k - x^*), \text{ thus } 0 = \nabla h_j^T(\xi^k) \frac{(x^k - x^*)}{\|(x^k - x^*)\|}.$$

In particular, it holds

$$0 = \lim_{l \to +\infty} \nabla h_j^T(\xi^{k_l}) \frac{(x^{k_l} - x^*)}{\|(x^{k_l} - x^*)\|} = \nabla h_j(x^*)^T d^*.$$

Now, let $i \in \mathcal{A}(x^*)$. By using again the Mean Value Theorem, for all $k \geq 1$ we know that there exists $\xi^k \in (x^*, x^k)$ such that, since $x^k \in X$, the following holds

$$0 \geq g_i(x^k) = g_i(x^*) + \nabla g_i(\xi^k)^T (x^k - x^*) = \nabla g_i(\xi^k)^T (x^k - x^*).$$

The same argument as above proves that $\nabla g_i(x^*)^T d^* \leq 0$.

Next, we will prove that $d^* \in T_2(x^*) \setminus \{0\}$. Since $\|d^*\| = 1$, we have that $d^* \neq 0$. Therefore, we only need to show that for all $i \in \mathcal{A}_>(x^*)$ it holds $\nabla g_i(x^*)^T d^* = 0$. Assume that there exists $\tilde{i} \in \mathcal{A}_>(x^*)$ it such that $\nabla g_{\tilde{i}}(x^*)^T d^* < 0$.

By using again the Mean Value Theorem, for all $k \geq 1$ we know that there exists $\xi^k \in (x^*, x^k)$ such that, by taking into account (3.2), the following holds

$$0 \geq f(x^k) - f(x^*) = \nabla f(\xi^k)^T (x^k - x^*).$$

By the same argument as above, we obtain that $\nabla f(x^*)^T d^* \leq 0$.

By using that (x^*, λ^*, μ^*) is a KKT point of (1.5), we have

$$\begin{aligned} 0 &\geq \nabla f(x^*)^T d^* \\ &= - \sum_{i \in \mathcal{A}_0(x^*)} \lambda_i^* \nabla g_i(x^*)^T d^* - \sum_{i \in \mathcal{A}_>(x^*)} \lambda_i^* \nabla g_i(x^*)^T d^* - \sum_{i \in I(x^*)} \lambda_i^* \nabla g_i(x^*)^T d^* - \sum_{j=1}^{p} \mu_j^* \nabla h_j(x^*)^T d^* \\ &= - \sum_{i \in \mathcal{A}_>(x^*)} \lambda_i^* \nabla g_i(x^*)^T d^* \geq \; -\lambda_{\tilde{i}}^* \nabla g_{\tilde{i}}(x^*)^T d^* > 0 \end{aligned}$$

which gives a contradiction. This demonstrates that $d^* \in T_2(x^*) \setminus \{0\}$.

Last, we will show that $(d^*)^T \nabla_{xx} L(x^*, \lambda^*, \mu^*) d^* \leq 0$. By Taylor's Theorem, for all $k \geq 1$ we know that there exists $\xi^k \in (x^*, x^k)$ such that

$$L(x^k, \lambda^*, \mu^*) = L(x^*, \lambda^*, \mu^*) + \nabla_x L(x^*, \lambda^*, \mu^*)^T (x^k - x^*) + \frac{1}{2} (x^k - x^*)^T \nabla_{xx}^2 L(\xi^k, \lambda^*, \mu^*)(x^k - x^*).$$

Since (x^*, λ^*, μ^*) is a KKT point of (1.5), it holds $\nabla_x L(x^*, \lambda^*, \mu^*) = 0$ and therefore for all $k \geq 1$ it holds

$$L(x^k, \lambda^*, \mu^*) = L(x^*, \lambda^*, \mu^*) + \frac{1}{2} (x^k - x^*)^T \nabla_{xx}^2 L(\xi^k, \lambda^*, \mu^*)(x^k - x^*). \tag{3.3}$$

Using again that (x^*, λ^*, μ^*) is a KKT point, we have $L(x^*, \lambda^*, \mu^*) = f(x^*)$, thus, by using (3.2), for all $k \geq 1$ it yields

$$L(x^*, \lambda^*, \mu^*) = f(x^*) \geq f(x^k) \geq f(x^k) + \sum_{i=1}^{m} \lambda_i^* g_i(x^k) + \sum_{j=1}^{p} \mu_j^* h(x^k) = L(x^k, \lambda^*, \mu^*).$$

Plugging this inequality into (3.3), we obtain that for all $k \geq 1$ it holds

$$0 \geq \frac{1}{2} \frac{(x^k - x^*)^T}{\|x^k - x^*\|} \nabla_{xx}^2 L(\xi^k, \lambda^*, \mu^*) \frac{x^k - x^*}{\|x^k - x^*\|}.$$

This allows us to conclude

$$0 \geq \lim_{l \to +\infty} \frac{1}{2} \frac{(x^{k_l} - x^*)^T}{\|x^{k_l} - x^*\|} \nabla_{xx} L(\xi^{k_l}, \lambda^*, \mu^*) \frac{x^{k_l} - x^*}{\|x^{k_l} - x^*\|} = \frac{1}{2} (d^*)^T \nabla_{xx}^2 L(x^*, \lambda^*, \mu^*) d^*,$$

which demonstrates that $(d^*)^T \nabla_{xx} L(x^*, \lambda^*, \mu^*) d^* \leq 0$. ■

Chapter II

Numerical methods for unconstrained optimization problems

4 A general descent algorithm

In the following, we will discuss a general descent algorithm for solving optimization problems of the form

$$\min_{x\in\mathbb{R}^n} f(x), \tag{4.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable function.

Definition 4.1 (descent direction) A vector $d \in \mathbb{R}^n$ is called a descent direction of f at $x \in \mathbb{R}^n$ if

$$\exists \bar{t} > 0 \text{ such that } f(x + td) < f(x) \ \ \forall\, t \in (0, \bar{t}]. \tag{4.2}$$

Lemma 4.2 *Let be $x \in \mathbb{R}^n$ and $d \in \mathbb{R}^n$ such that $\nabla f(x)^T d < 0$. Then d is a descent direction of f at x.*

Proof. By the definition of directional derivative, we have that

$$\lim_{t\to 0} \frac{f(x + td) - f(x)}{t} = \nabla f(x)^T d < 0.$$

Therefore, there exists $\bar{t} > 0$ such that, for every $t \in (0, \bar{t}]$, we have

$$\frac{f(x + td) - f(x)}{t} < 0 \Leftrightarrow f(x + td) < f(x).$$

■

Remark 4.3 Asking for $\nabla f(x)^T d < 0$ is equivalent to asking for the angle between $-\nabla f(x)$ and d to be acute. This is because we have

$$\cos(\sphericalangle(-\nabla f(x), d)) = -\frac{\nabla f(x)^T d}{\|\nabla f(x)\| \, \|d\|} > 0.$$

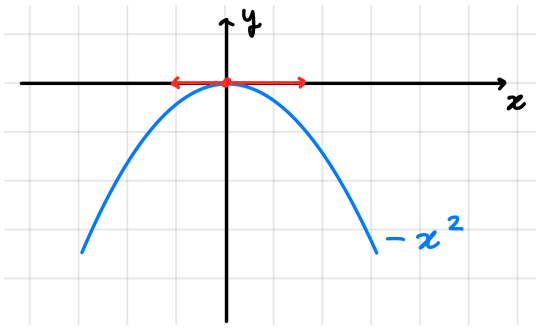

Figure 4.1: Every $d \in \mathbb{R} \setminus \{0\}$ is a descent direction of f at $x = 0$.

Example 4.4 The condition $\nabla f(x)^T d < 0$ is not necessary for d being a descent direction of f at x. Take, for example, $f(x) = -x^2$ and $x = 0$. Then every $d \in \mathbb{R} \setminus \{0\}$ is a descent direction of f at $x = 0$ (see Figure 4.1), but $f'(0)d = 0$.

Remark 4.5 Assume that x is not a critical point of f, meaning $\nabla f(x) \neq 0$. Then $d := -\nabla f(x)$ is a descent direction of f at x, since

$$\nabla f(x)(-\nabla f(x)) = -\|\nabla f(x)\|^2 < 0.$$

Moreover, if $B \in \mathbb{R}^{n \times n}$ is a symmetric positive definite matrix, then $d := -B\nabla f(x) \in \mathbb{R}^n$ is also a descent direction of f at x, since

$$\nabla f(x)^T(-B\nabla f(x)) = -\nabla f(x)^T B \nabla f(x) < 0.$$

Algorithm 4.6 (general line search algorithm)

1: Choose a starting point $x^0 \in \mathbb{R}^n$ and set $k := 0$.

2: If x^k fulfills a stopping criterion: STOP.

3: Find a descent direction d^k of f at x^k.

4: Find a step size $t_k > 0$ such that $f(x^k + t_k d^k) < f(x^k)$.

5: Set $x^{k+1} := x^k + t_k d^k$, $k := k + 1$ and go to Step 2.

The line search algorithm has two degrees of freedom: the choice of the descent direction (Step 3) and the choice of the step size (Step 4).

Example 4.7 The step size cannot be chosen arbitrarily. The function $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$, has $x^* = 0$ as its unique global minimum and $d = -1$ is a descent direction of f at every $x > 0$. We choose $x^0 := 1$ and $t_k := \frac{1}{2^{k+2}}$ for every $k \geq 0$. Then it holds every $k \geq 0$

$$x^{k+1} = x^k - t_k = \ldots = x^0 - t_0 - t_1 - \ldots - t_k = 1 - \sum_{i=0}^{k} \frac{1}{2^{i+2}},$$

therefore $x^k \to \frac{1}{2}$ as $k \to +\infty$, which is not the global minimum of f.

Definition 4.8 (step size strategy)

(a) A set-valued mapping $T : \mathbb{R}^n \times \mathbb{R}^n \rightrightarrows (0, +\infty)$ which assigns to each pair $(x, d) \in \mathbb{R}^n \times \mathbb{R}^n$ a set of step sizes $T(x, d) \subseteq (0, +\infty)$ is called step size strategy.

(b) A step size strategy $T : \mathbb{R}^n \times \mathbb{R}^n \rightrightarrows (0, +\infty)$ is called well-defined if, for every $(x, d) \in \mathbb{R}^n \times \mathbb{R}^n$ fulfilling $\nabla f(x)^T d < 0$, it holds that $T(x, d) \neq \emptyset$.

(c) A step size strategy $T : \mathbb{R}^n \times \mathbb{R}^n \rightrightarrows (0, +\infty)$ is called efficient if there exists $\theta > 0$ such that, for every $(x, d) \in \mathbb{R}^n \times \mathbb{R}^n$ where d is a descent direction of f at x, it holds

$$f(x + td) \leq f(x) - \theta \left(\frac{\nabla f(x)^T d}{\|d\|} \right)^2 \quad \forall t \in T(x, d).$$

In this case, every step size $t \in T(x, d)$ is called efficient.

In the following we assume that Algorithm 4.6 does not terminate after finitely many iterations, which means that it generates an infinite sequence of iterates $(x^k)_{k \geq 0}$.

Theorem 4.9 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable and $(x^k)_{k \geq 0}$ a sequence generated by Algorithm 4.6 such that*

(a) the so-called angle condition holds, i.e.

$$\exists c > 0 \text{ such that } \frac{-\nabla f(x^k)^T d^k}{\|\nabla f(x^k)\| \, \|d^k\|} \geq c \quad \forall k \geq 0;$$

(b) every step size $t_k \in T(x^k, d^k), k \geq 0$, is efficient.

Then every accumulation (limit) point of $(x^k)_{k \geq 0}$ is a critical point of f.

Proof. Since every step size is efficient, we have for every $k \geq 0$

$$f(x^{k+1}) = f(x^k + t_k d^k) \leq f(x^k) - \theta \left(\frac{\nabla f(x^k)^T d^k}{\|d^k\|} \right)^2$$

and, by the angle condition,

$$\left(\frac{\nabla f(x^k)^T d^k}{\|d^k\|} \right)^2 \geq c^2 \left\| \nabla f(x^k) \right\|^2.$$

Combining these two inequalities, we get for every $k \geq 0$

$$f(x^{k+1}) \leq f(x^k) - c^2 \theta \left\| \nabla f(x^k) \right\|^2 \leq f(x^k), \tag{4.3}$$

which implies that the sequence $(f(x^k))_{k \geq 0}$ is nonincreasing.

Let us choose an accumulation point $x^* \in \mathbb{R}^n$ of the sequence $(x^k)_{k\geq 0}$, meaning that there exists a subsequence $(x^{k_l})_{l\geq 0)}$ such that $x^{k_l} \to x^*$ as $l \to +\infty$. By the continuity of f, we have that $f(x^{k_l}) \to f(x^*)$ as $l \to +\infty$. Combining this last observation with the fact that $(f(x^k))_{k\geq 0}$ is nonincreasing, we can conclude that

$$f(x^k) \to f(x^*) \quad (k \to +\infty).$$

From (4.3), it yields for every $k \geq 0$

$$0 \leq c^2\theta \left\|\nabla f(x^k)\right\|^2 \leq f(x^k) - f(x^{k+1}).$$

Since the right-hand side converges to zero as $k \to +\infty$, we get $\left\|\nabla f(x^k)\right\| \to 0$ as $k \to +\infty$. By continuity of the gradient we know that $\left\|\nabla f(x^{k_l})\right\| \to \|\nabla f(x^*)\|$ as $l \to +\infty$, and this allows us to conclude that $\nabla f(x^*) = 0$. ■

Remark 4.10 (a) The angle condition states that the angle $\sphericalangle(-\nabla f(x^k), d^k)$ stays uniformly away from 90°. It is for instance fulfilled for $d^k := -B\nabla f(x^k)$ for every $k \geq 0$, where $B \in \mathbb{R}^{n\times n}$ is a symmetric and positive definite matrix, and $0 < c \leq \frac{\lambda_{\min}(B)}{\|B\|}$.

(b) If f is coercive, namely $\lim_{\|x\|\to+\infty} f(x) = +\infty$, then the lower level set of f at x^0,

$$\mathcal{L}(x^0) := \{x \in \mathbb{R}^n : f(x) \leq f(x^0)\},$$

which contains according to (4.3) the entire sequence $(x^k)_{k\geq 0}$, is bounded. In this case, $(x^k)_{k\geq 0}$ has an accumulation point.

(c) How to choose an "optimal" step size? One could, for example, consider the minimization rule, which consists in choosing $t := t_{\min}$ such that

$$f(x + t_{\min}d) = \min_{t>0} f(x + td).$$

This rule is, under certain assumptions, well-defined and efficient. However, the step size cannot always be calculated explicitly. One exception is for

$$f : \mathbb{R}^n \to \mathbb{R} \ , \ f(x) = \frac{1}{2}x^T Ax - b^T x,$$

with $A \in \mathbb{R}^{n\times n}$ a symmetric and positive definite matrix and $b \in \mathbb{R}^n$. Indeed, for $x, d \in \mathbb{R}^n$ with $\nabla f(x)^T d < 0$, the step size defined by the minimization rule is well-defined and efficient and it can be explicitly calculated, namely,

$$t_{\min} = -\frac{\nabla f(x)^T d}{d^T Ad}.$$

5 Step size strategies

In this section, we will discuss three "popular" step size strategies used for the minimization of a continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$. Throughout this section we will assume that

$$x, d \in \mathbb{R}^n \text{ are given such that } \nabla f(x)^T d < 0. \tag{5.1}$$

5.1 The Wolfe-Powell step size strategy

Let $\sigma \in (0, \frac{1}{2})$ and $\rho \in [\sigma, 1)$. The Wolfe-Powell step size strategy consists of finding $t^* > 0$ such that

$$f(x + t^* d) \leq f(x) + \sigma t^* \nabla f(x)^T d \tag{5.2}$$

and

$$\nabla f(x + t^* d)^T d \geq \rho \nabla f(x)^T d. \tag{5.3}$$

We define $\phi : \mathbb{R} \to \mathbb{R}, \phi(t) := f(x + td)$. It follows that $\phi'(t) = \nabla f(x + td)^T d$ for all $t \in \mathbb{R}$ and $\phi'(0) = \nabla f(x)^T d < 0$. We can write the conditions above as follows

$$(5.2) \;\Leftrightarrow\; \phi(t^*) \leq \phi(0) + \sigma t^* \phi'(0)$$

and

$$(5.3) \;\Leftrightarrow\; \phi'(t^*) \geq \rho \phi'(0).$$

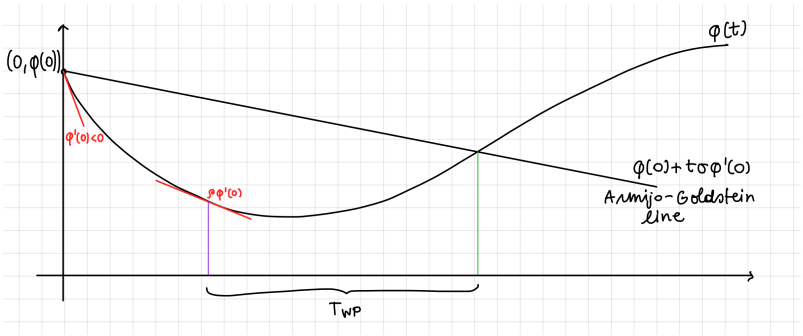

Figure 5.1: Geometric interpretation of the Wolfe-Powell step size strategy for $\varphi(x) = f(x + td)$

In other words, the step size $t^* > 0$ is chosen such that the following two conditions are satisfied:

- the graph of ϕ at t^* lies below the Armijo-Goldstein line;
- the graph of ϕ at t^* decreases less steeply than it does at 0 or even increases.

Definition 5.1 (Wolfe-Powell step size strategy) Let $\sigma \in (0, \frac{1}{2})$, $\rho \in [\sigma, 1)$ and $x^0 \in \mathbb{R}^n$. For $x \in \mathcal{L}(x^0)$ and $d \in \mathbb{R}^n$ such that $\nabla f(x)^T d < 0$, we define

$$T_{WP}(x, d) := \left\{ t > 0 : f(x + td) \leq f(x) + \sigma t \nabla f(x)^T d \text{ and } \nabla f(x + td)^T d \geq \rho \nabla f(x)^T d \right\}$$

as the set of the Wolfe-Powell step size strategies in x in direction d.

Theorem 5.2 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable, $\sigma \in (0, \frac{1}{2})$, $\rho \in [\sigma, 1)$ and $x^0 \in \mathbb{R}^n$. For $x \in \mathcal{L}(x^0)$ and $d \in \mathbb{R}^n$ such that $\nabla f(x)^T d < 0$, let $T_{WP}(x, d)$ be the set of Wolfe-Powell step size strategies in x in direction d. Then the following statements are true:*

(a) If f is bounded from below, then $T_{WP}(x, d)$ is nonempty. In other words, the step size strategy is well-defined.

(b) If ∇f is Lipschitz continuous on $\mathcal{L}(x^0)$, then there exists $\theta > 0$ such that

$$f(x + td) \leq f(x) - \theta \left(\frac{\nabla f(x)^T d}{\|d\|} \right)^2 \text{ for all } t \in T_{WP}(x, d).$$

In other words, the step size strategy is efficient.

Proof. For all $t \in \mathbb{R}$, we define

$$\phi(t) = f(x + td) \quad \text{and} \quad \psi(t) = f(x) + \sigma t \nabla f(x)^T d. \tag{5.4}$$

(a) It suffices to show that there exists $t^* > 0$ such that

$$\phi(t^*) \leq \psi(t^*) \quad \text{and} \quad \phi'(t^*) \geq \rho \phi'(0).$$

Since $\sigma < \frac{1}{2} < 1$ and $\nabla f(x)^T d < 0$, we have

$$\phi'(0) = \nabla f(x)^T d < \sigma \nabla f(x)^T d = \psi'(0).$$

Therefore, $(\psi - \phi)'(0) > 0$ and, by definition of the derivative,

$$\lim_{t \downarrow 0} \frac{(\psi - \phi)(t)}{t} = \lim_{t \downarrow 0} \frac{(\psi - \phi)(t) - (\psi - \phi)(0)}{t} > 0.$$

Therefore, there exists $t_0 > 0$ such that, for all $t \in (0, t_0)$, it holds $\psi(t) > \phi(t)$. We choose t^* as being the first $t > 0$ at which ϕ and ψ intersect, namely,

$$t^* := \min\{t > 0 : \phi(t) = \psi(t)\}.$$

Note that t^* exists, since $\lim_{t \to \infty} \psi(t) = -\infty$ and ϕ is bounded from below, which is due to the fact that f is bounded from below. Then we have

$$\phi'(t^*) = \lim_{\substack{t \to t^* \\ t < t^*}} \frac{\phi(t) - \phi(t^*)}{t - t^*} \geq \lim_{\substack{t \to t^* \\ t < t^*}} \frac{\psi(t) - \psi(t^*)}{t - t^*} = \psi'(t^*)$$

where the inequality above holds since $\phi(t^*) = \psi(t^*)$, $\phi(t) < \psi(t)$ for $t < t^*$ and $t - t^* < 0$. In addition,

$$\psi'(t^*) = \sigma \nabla f(x)^T d \geq \rho \nabla f(x)^T d = \rho \phi'(0).$$

In conclusion, $t^* \in T_{WP}(x, d)$.

(b) Let $t \in T_{WP}(x,d)$. it holds

$$\phi(t) = f(x+td) \leq \psi(t) = f(x) + t\sigma\nabla f(x)^T d < f(x) \leq f(x^0),$$

which implies that $x + td \in \mathcal{L}(x^0)$. Furthermore, since $\rho < 1$ and $\nabla f(x)^T d < 0$, it holds

$$\begin{aligned} 0 &< (\rho - 1)\nabla f(x)^T d = \rho\nabla f(x)^T d - \nabla f(x)^T d \\ &\leq \nabla f(x+td)^T d - \nabla f(x)^T d = (\nabla f(x+td) - \nabla f(x))^T d \\ &\leq \|\nabla f(x+td) - \nabla f(x)\| \, \|d\| \leq Lt \, \|d\|^2 , \end{aligned}$$

where $L > 0$ denotes the Lipschitz constant of ∇f on $\mathcal{L}(x^0)$.

Defining

$$\theta := \frac{\sigma(1-\rho)}{L},$$

we obtain

$$f(x+td) \leq f(x) + \sigma t \nabla f(x)^T d \leq f(x) + \frac{\sigma(\rho-1)}{L}\left(\frac{\nabla f(x)^T d}{\|d\|}\right)^2 = f(x) - \theta\left(\frac{\nabla f(x)^T d}{\|d\|}\right)^2.$$

■

Remark 5.3 Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable and bounded from below, $\sigma \in (0, \frac{1}{2})$, $\rho \in (\sigma, 1)$, $x, d \in \mathbb{R}^n$ such that $\nabla f(x)^T d < 0$, and

$$\phi(t) = f(x+td) \quad \text{and} \quad \psi(t) = f(x) + \sigma t \nabla f(x)^T d.$$

We consider the following algorithm:
Phase A:

A0: Choose $t_0 > 0$ and set $k := 0$.

A1: If $\phi(t_k) \geq \psi(t_k)$, then set $a := 0$ and $b := t_k$ and go to Step B0.

If $\phi(t_k) < \psi(t_k)$ and $\phi'(t_k) \geq \rho\phi'(0)$, then set $t^* := t_k$ and terminate: STOP 1.

If $\phi(t_k) < \psi(t_k)$ and $\phi'(t_k) < \rho\phi'(0)$, then set $t_{k+1} := 2t_k$, $k := k+1$ and go to Step A1.

Phase B:

B0: Set $k := 0$, and adopt $a_0 := a$ and $b_0 := b$ from Phase A.

B1: Set $t_k := \frac{a_k + b_k}{2}$.

B2: If $\phi(t_k) \geq \psi(t_k)$, then set $a_{k+1} := a_k$, $b_{k+1} := t_k$, $k := k+1$, and go to Step B1.

If $\phi(t_k) < \psi(t_k)$ and $\phi'(t_k) \geq \rho\phi'(0)$, then set $t^* := t_k$ and terminate: STOP 2.

If $\phi(t_k) < \psi(t_k)$ and $\phi'(t_k) < \rho\phi'(0)$, then set $a_{k+1} := t_k$, $b_{k+1} := b_k$, $k := k+1$, and go to Step B1.

The algorithm terminates after a finite number of steps at either STOP 1 or STOP 1, providing a Wolfe–Powell step size t^* (see [4]).

5.2 The strong Wolfe-Powell step size strategy

In this section, we consider a step size strategy that refines the Wolfe-Powell step size strategy by also bounding the increase of ϕ from above.

Let $\sigma \in (0, \frac{1}{2})$ and $\rho \in [\sigma, 1]$. The strong Wolfe-Powell step size strategy consists of finding $t^* > 0$ such that

$$f(x + t^* d) \leq f(x) + \sigma t^* \nabla f(x)^T d \tag{5.5}$$

and

$$|\nabla f(x + t^* d)^T d| \leq -\rho \nabla f(x)^T d. \tag{5.6}$$

We can write the conditions above as follows

$$(5.5) \;\Leftrightarrow\; \phi(t^*) \leq \phi(0) + \sigma t^* \phi'(0)$$

and

$$(5.6) \;\Leftrightarrow\; |\phi'(t^*)| \leq -\rho \phi'(0).$$

In other words, the step size $t^* > 0$ is chosen such that the following two conditions are satisfied:

- the graph of ϕ at t^* lies below the Armijo-Goldstein line;
- the graph of ϕ at t^* either decreases or increases less steeply than it decreases at 0.

Definition 5.4 (strong Wolfe-Powell step size strategy) Let $\sigma \in (0, \frac{1}{2})$, $\rho \in [\sigma, 1)$ and $x^0 \in \mathbb{R}^n$. For $x \in \mathcal{L}(x^0)$ and $d \in \mathbb{R}^n$ such that $\nabla f(x)^T d < 0$, we define

$$T_{SWP}(x, d) := \left\{ t > 0 : f(x + td) \leq f(x) + \sigma t \nabla f(x)^T d \text{ and } |\nabla f(x + td)^T d| \leq -\rho \nabla f(x)^T d \right\}$$

as the set of the strong Wolfe-Powell step size strategies in x in direction d.

Theorem 5.5 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable, $\sigma \in (0, \frac{1}{2})$, $\rho \in [\sigma, 1)$ and $x^0 \in \mathbb{R}^n$. For $x \in \mathcal{L}(x^0)$ and $d \in \mathbb{R}^n$ such that $\nabla f(x)^T d < 0$, let $T_{WP}(x, d)$ be the set of strong Wolfe-Powell step size strategies in x in direction d. Then the following statements are true:*

(a) If f is bounded from below, then $T_{SWP}(x, d)$ is nonempty. In other words, the step size strategy is well-defined.

(b) If ∇f is Lipschitz continuous on $\mathcal{L}(x^0)$, then there exists $\theta > 0$ such that

$$f(x + td) \leq f(x) - \theta \left(\frac{\nabla f(x)^T d}{\|d\|} \right)^2 \text{ for all } t \in T_{SWP}(x, d).$$

In other words, the step size strategy is efficient.

Proof. As in the proof of Theorem 5.2, we denote

$$\phi(t) = f(x + td) \quad \text{and} \quad \psi(t) = f(x) + \sigma t \nabla f(x)^T d.$$

(a) As in the proof of Theorem 5.2 (a), there exists

$$t^* := \min\{t > 0 : \phi(t) = \psi(t)\}$$

with $\phi'(t^*) \geq \psi'(t^*)$.

First, we consider the case when $\phi'(t^*) \leq 0$. We have

$$|\nabla f(x + t^* d)^T d| = |\phi'(t^*)| = -\phi'(t^*) \leq -\psi'(t^*) = -\sigma\phi'(0) \leq -\rho\phi'(0),$$

therefore, $t^* \in T_{SWP}(x, d)$.

Now, assume $\phi'(t^*) > 0$. Since $\nabla f(x)^T d = \phi'(0) < 0$, there exists $t^{**} \in (0, t^*)$ such that $\phi'(t^{**}) = 0$. Therefore,

$$\phi(t^{**}) < \psi(t^{**}) \Leftrightarrow f(x + t^{**} d) < f(x) + \sigma t^{**} \nabla f(x)^T d.$$

Furthermore,

$$|\nabla f(x + t^{**} d)^T d| = |\phi'(t^{**})| = 0 \leq -\rho\phi'(0) = -\phi \nabla f(x)^T d,$$

which implies $t^{**} \in T_{SWP}(x, d)$.

(b) Follows from Theorem 5.2 (b), since $T_{SWP}(x, d) \subseteq T_{WP}(x, d)$. ■

Remark 5.6 As for the Wolfe-Powell strategy, it is possible to construct an algorithm that determines a strong Wolfe-Powell step size in a finite number of steps.

5.3 The (backtracking) Armijo rule

In the following section, we will introduce an easy-to-implement step-size strategy that is not efficient.

Let $\sigma \in (0, 1)$, $\beta \in (0, 1)$ and $x, d \in \mathbb{R}^n$ such that $\nabla f(x)^T d < 0$. The Armijo step size consists of choosing

$$t := \max\{\beta^l : l = 0, 1, 2, \ldots\} \text{ such that } f(x + td) \leq f(x) + \sigma t \nabla f(x)^T d.$$

Theorem 5.7 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable, $\sigma \in (0, 1)$ and $\beta \in (0, 1)$. For every $x, d \in \mathbb{R}^n$ such that $\nabla f(x)^T d < 0$ there exists $l \geq 0$ such that*

$$f(x + \beta^l d) \leq f(x) + \sigma \beta^l \nabla f(x)^T d.$$

In other words, the Armijo rule is well-defined.

Proof. Let $x, d \in \mathbb{R}^n$ such that $\nabla f(x)^T d < 0$ and assume that for all $l \geq 0$

$$f(x + \beta^l d) > f(x) + \sigma \beta^l \nabla f(x)^T d$$

or, equivalently,

$$\frac{f(x + \beta^l d) - f(x)}{\beta^l} > \sigma \nabla f(x)^T d.$$

We let $l \to +\infty$, and obtain $\nabla f(x)^T d \geq \sigma \nabla f(x)^T d$, which leads to the desired contradiction. ■

Remark 5.8 The Armijo rule is not efficient. For the function $f : \mathbb{R} \to \mathbb{R}$, $f(x) = \frac{x^2}{2}$, we choose $x^0 = -3, \sigma = \frac{1}{2}$, and, for every $k \geq 0$, $d^k := \frac{1}{2^k}$ and $x^{k+1} := x^k + t_k d^k$, where t_k is the Armijo step size.

It holds that $t_k = 1$ for every $k \geq 0$ and $x^k = -3 + 1 + \frac{1}{2} + \ldots + \frac{1}{2^{k-1}} = -1 - \frac{1}{2^{k-1}}$ for every $k \geq 1$. Indeed, we have for $xd < 0$

$$\frac{(x + td)^2}{2} \leq \frac{x^2}{2} + \frac{1}{2} txd \;\Leftrightarrow\; t^2 d^2 \leq -txd \;\Leftrightarrow\; t \leq -\frac{x}{d}.$$

Carrying out the calculations for $k = 0$ and $k = 1$, we obtain, respectively,

$$x^0 = -3,\ d^0 = 1 :\ -\frac{x^0}{d^0} = 3, l = 0 \text{ and } t_0 = \beta^0 = 1$$
$$x^1 = -3 + 1 = -2,\ d^1 = \frac{1}{2} :\ -\frac{x^1}{d^1} = 4, l = 0 \text{ and } t_1 = \beta^0 = 1.$$

and, by induction,

$$x^k = -1 - \frac{1}{2^{k-1}},\ d^k = \frac{1}{2^k} :\ -\frac{x^k}{d^k} = 2^k + 2, l = 0 \text{ and } t_k = \beta^0 = 1.$$

Furthermore, it can be shown by contradiction that there exists no $\theta > 0$ such that for every $k \geq 0$ (notice that $x^k d^k < 0$) and every $t = 1$ we have

$$\frac{(x^k + td^k)^2}{2} \leq \frac{(x^k)^2}{2} - \theta \left(\frac{x^k d^k}{d^k} \right)^2 = \frac{(x^k)^2}{2} - \theta (x^k)^2,$$

which is equivalent to

$$\theta \leq -\frac{d^k}{x^k} - \frac{1}{2} \left(\frac{d^k}{x^k} \right)^2 = \frac{1}{2^k + 2} - \frac{1}{2(2^k + 2)^2}.$$

The contradiction is obtained as $k \to +\infty$.

6 The gradient algorithm

In this section we will introduce and analyze the gradient algorithm with Armijo step size rule for the minimization of a continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$.

Algorithm 6.1 (gradient algorithm with Armijo step size rule)

1: Choose a starting point $x^0 \in \mathbb{R}^n$, $\sigma \in (0,1)$, $\beta \in (0,1)$, $\varepsilon \geq 0$ and set $k := 0$.

2: If $\left\|\nabla f(x^k)\right\| \leq \varepsilon$: STOP.

3: Set $d^k := -\nabla f(x^k)$.

4: Find the Armijo step size

$$t_k := \max\{\beta^l : l = 0, 1, \ldots\}$$

with

$$f(x^k + t_k d^k) \leq f(x^k) + \sigma t_k \nabla f(x^k)^T d^k.$$

5: Set $x^{k+1} := x^k + t_k d^k$, $k := k + 1$ and go to Step 2.

For the analysis of the gradient algorithm, we will assume that $\varepsilon = 0$ and that Algorithm 6.1 does not terminate after finitely many steps.

The following lemma will play an important role in the proof of the main convergence theorem.

Lemma 6.2 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable functon, $(x^k)_{k\geq 0} \subseteq \mathbb{R}^n$, $(d^k)_{k\geq 0} \subseteq \mathbb{R}^n$ and $(t^k)_{k\geq 0} \subseteq \mathbb{R}$ such that $x^k \to x \in \mathbb{R}^n$, $d^k \to d \in \mathbb{R}^n$ and $t_k \downarrow 0$ as $k \to +\infty$. Then it holds*

$$\lim_{k\to+\infty} \frac{f(x^k + t_k d^k) - f(x^k)}{t_k} = \nabla f(x)^T d.$$

Proof. Let $k \geq 0$. According to the Mean Value Theorem, there exists $\xi^k \in (x^k, x^k + t_k d^k)$ such that

$$f(x^k + t_k d^k) - f(x^k) = \nabla f(\xi^k)^T (t_k d^k) \Leftrightarrow \frac{f(x^k + t_k d^k) - f(x^k)}{t_k} = \nabla f(\xi^k)^T (d^k).$$

The conclusion follows from the continuity of the gradient and the fact that $\xi^k \to x$ as $k \to +\infty$. ■

Theorem 6.3 (convergence of the gradient method) *Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable. Then every accumulation (limit) point of the sequence $(x^k)_{k\geq 0}$ generated by Algorithm 6.1 is a critical point of f.*

Proof. Let $x^* \in \mathbb{R}^n$ be a limit point of $(x^k)_{k\geq 0}$, i.e., there exists a subsequence $(x^{k_l})_{l\geq 0}$ such that $x^{k_l} \to x^*$ as $l \to +\infty$. Assume that x^* is not a critical point of f, i.e., $\nabla f(x^*) \neq 0$. For every $k \geq 0$ it holds

$$f(x^{k+1}) = f(x^k + t_k d^k) \leq f(x^k) + \sigma t_k \nabla f(x^k)^T d^k = f(x^k) - \sigma t_k \left\|\nabla f(x^k)\right\|^2 \leq f(x^k),$$

therefore, $(f(x^k))_{k\geq 0}$ is nonincreasing.

Since f is continuous, we have $f(x^{k_l}) \to f(x^*)$ as $l \to +\infty$ and, therefore, $f(x^k) \to f(x^*)$ as $k \to +\infty$. Since for every $k \geq 0$

$$0 \leq \sigma t_k \left\|\nabla f(x^k)\right\|^2 \leq f(x^k) - f(x^{k+1})$$

and $f(x^k) - f(x^{k+1}) \to 0$ as $k \to +\infty$, it yields $t_k \left\|\nabla f(x^k)\right\|^2 \to 0$ as $k \to +\infty$, and so $t_{k_l} \left\|\nabla f(x^{k_l})\right\|^2 \to 0$ as $l \to +\infty$. Since $\|\nabla f(x^*)\| \neq 0$, we can conclude that $t_{k_l} \to 0$ as $l \to +\infty$.

For every $l \geq 0$, we have $t_{k_l} := \beta^{m_{k_l}}$. By the Armijo rule, it holds for every $l \geq 0$

$$f(x^{k_l} + \beta^{m_{k_l}-1} d^{k_l}) > f(x^{k_l}) + \sigma \beta^{m_{k_l}-1} \nabla f(x^{k_l})^T d^{k_l}$$

or, equivalently,

$$\frac{f(x^{k_l} + \beta^{m_{k_l}-1} d^{k_l}) - f(x^{k_l})}{\beta^{m_{k_l}-1}} > \sigma \nabla f(x^{k_l})^T d^{k_l}.$$

Note that this holds because, by the Armijo rule, m_{k_l} is the first exponent for which the inequality $f(x^{k_l}+\beta^{m_{k_l}} d^{k_l}) \leq f(x^{k_l})+\sigma\beta^{m_{k_l}} \nabla f(x^{k_l})^T d^{k_l}$ is fulfilled, so for $\beta^{m_{k_l}-1}$, the inequality is not fulfilled.

Observe that

$$\begin{aligned}
&\beta^{m_{k_l}-1} = \frac{1}{\beta} t_{k_l} \to 0 \ \ (l \to +\infty),\\
&x^{k_l} \to x^* \ \ (l \to +\infty),\\
&d^{k_l} = -\nabla f(x^{k_l}) \to -\nabla f(x^*) \ \ (l \to +\infty),
\end{aligned}$$

which, according to Lemma 6.2, gives

$$\nabla f(x^*)^T(-\nabla f(x^*)) \geq \sigma \nabla f(x^*)^T(-\nabla f(x^*)) \ \Leftrightarrow \ -\left\|\nabla f(x^*)\right\|^2 \geq -\sigma \left\|\nabla f(x^*)\right\|^2,$$

which is a contradiction to $\sigma \in (0,1)$. Thus, $\nabla f(x^*) = 0$. ■

7 The gradient method for convex optimization problems

In this section, we will analyze the convergence properties of the gradient method for solving the unconstrained minimization problem

$$\min_{x\in\mathbb{R}^n} f(x) \tag{7.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a convex and differentiable function with a $L_{\nabla f}$-Lipschitz continuous gradient, meaning that there exists $L_{\nabla f} > 0$ such that

$$\|\nabla f(x) - \nabla f(y)\| \leq L_{\nabla f} \|x - y\| \quad \forall x, y \in \mathbb{R}^n.$$

Throughout this section we will assume that the set of (global) minimizers of f

$$\operatorname{argmin} f := \{x^* \in \mathbb{R}^n : f(x^*) = \inf_{x \in \mathbb{R}^n} f(x)\}$$

is nonempty, therefore

$$f_* := \inf_{x \in \mathbb{R}^n} f(x) \in \mathbb{R}.$$

7.1 Gradient flow

In this subsection, we will investigate the asymptotic properties of the following so-called gradient flow system

$$\begin{cases} \dot{x}(t) = -\nabla f(x(t)) \\ x(t_0) = x^0 \in \mathbb{R}^n \end{cases} \tag{7.2}$$

defined on $[t_0, +\infty)$, for $t_0 \geq 0$, that we attach to the convex minimization problem (7.1).

For every $t_0 \geq 0$ and every $x^0 \in \mathbb{R}^n$, the global version of the Cauchy-Lipschitz Theorem guarantees the existence and uniqueness of a solution $x \in C^1([t_0, +\infty))$ of (7.2).

The existence and uniqueness of a solution $x \in C^1([t_0, +\infty))$ of (7.2) can be guaranteed by only assuming that ∇f is Lipschitz continuous on every bounded set of $\mathbb{R}^n$. To this end one could make use of the local version of the Cauchy-Lipschitz Theorem together with maximal-time arguments.

Before discussing the asymptotic behaviour of the gradient flow, we will prove the following useful technical result.

Lemma 7.1 *Let $F : [t_0, +\infty) \to \mathbb{R}$ be a locally absolutely continuous and bounded from below function and $G : [t_0, +\infty) \to \mathbb{R}$ an L^1-integrable function such that*

$$\frac{d}{dt} F(t) \leq G(t) \quad \text{for almost every} \quad t \geq t_0.$$

Then there exists $\lim_{t \to +\infty} F(t) \in \mathbb{R}$.

Proof. Using that $\int_{t_0}^{+\infty} \max\{G(u), 0\} du \leq \int_{t_0}^{+\infty} |G(u)| du < +\infty$ and $\int_{t_0}^{+\infty} \max\{-G(u), 0\} du \leq \int_{t_0}^{+\infty} |G(u)| du < +\infty$, we have

$$\lim_{t \to +\infty} \int_{t_0}^{t} G(u) du = \lim_{t \to +\infty} \int_{t_0}^{t} (\max\{G(u), 0\} - \max\{-G(u), 0\}) du \text{ exists and is finite.}$$

Let $t_0 \leq s \leq t$. By integration, we obtain

$$F(t) - F(s) \leq \int_s^t G(u) du$$

or, equivalently,

$$F(t) - \int_{t_0}^{t} G(u)du \leq F(s) - \int_{t_0}^{s} G(u)du.$$

This means that the function $t \mapsto F(t) - \int_{t_0}^{t} G(u)du$ is nonincreasing and bounded from below. Therefore, the limit $\lim_{t\to+\infty} F(t) - \int_{t_0}^{t} G(u)du$ exists and is finite., which implies that $\lim_{t\to+\infty} F(t)$ exists and is finite. ■

Theorem 7.2 (asymptotic behaviour of the gradient flow) *Let $x : [t_0, +\infty) \to \mathbb{R}^n$ be a trajectory solution of the gradient flow system (7.2). Then the following statements are true:*

(a) it holds

$$\int_{t_0}^{+\infty} t\|\dot{x}(t)\|^2 dt < +\infty, \int_{t_0}^{+\infty} t\|\nabla f(x(t))\|^2 dt < +\infty \quad \text{and} \quad \int_{t_0}^{+\infty} (f(x(t)) - f_*)dt < +\infty;$$

(b) it holds $f(x(t)) - f_* = o\left(\frac{1}{t}\right)$ as $t \to +\infty$;

(c) $x(t)$ converges to an element in argmin f *as $t \to +\infty$;*

(d) it holds $\|\nabla f(x(t))\| = o\left(\frac{1}{t}\right)$ as $t \to +\infty$;

Proof. For $x^* \in \operatorname{argmin} f$, we define the energy function

$$\mathcal{E} : [t_0, +\infty) \to \mathbb{R}, \quad \mathcal{E}(t) = t(f(x(t)) - f_*) + \frac{1}{2}\|x(t) - x_*\|^2.$$

For all $t \geq t_0$, it holds

$$\begin{aligned}\frac{d}{dt}\mathcal{E}(t) =& f(x(t)) - f_* + t\nabla f(x(t))^T\dot{x}(t) + (x(t) - x_*)^T\dot{x}(t)\\ =& f(x(t)) - f_* - t\|\dot{x}(t)\|^2 + (x(t) - x_*)^T\nabla f(x(t)) \leq -t\|\dot{x}(t)\|^2.\end{aligned}$$

In the last estimate, we used the gradient inequality.

By integration, for all $t \geq t_0$, it yields

$$\begin{aligned}& t(f(x(t)) - f_*) + \frac{1}{2}\|x(t) - x_*\|^2 + \int_{t_0}^{t} s\|\dot{x}(s)\|^2 ds\\ =& \mathcal{E}(t) + \int_{t_0}^{t} s\|\dot{x}(s)\|^2 ds\\ \leq& \mathcal{E}(t_0) = t_0(f(x(t_0)) - f_*) + \frac{1}{2}\|x(t_0) - x_*\|^2 = t_0(f(x^0) - f_*) + \frac{1}{2}\|x^0 - x_*\|^2.\end{aligned}$$

This yields that $x(\cdot)$ is bounded, $\int_{t_0}^{+\infty} t\|\dot{x}(t)\|^2 dt = \int_{t_0}^{+\infty} t\|\nabla f(x(t))\|^2 dt < +\infty$, and

$$0 \leq f(x(t)) - f_* \leq \frac{t_0(f(x^0) - f_*) + \frac{1}{2}\|x^0 - x_*\|^2}{t} \quad \forall t \geq t_0.$$

In other words, $f(x(t)) - f_* = \mathcal{O}\left(\frac{1}{t}\right)$ as $t \to +\infty$.

On the other hand, by using again the gradient inequality, it holds

$$\frac{d}{dt}\left(\frac{1}{2}\|x(t) - x_*\|^2\right) = (x(t) - x_*)^T \dot{x}(t) = -\nabla f(x(t))^T (x(t) - x_*) \leq -\left(f(x(t)) - f_*\right) \leq 0 \ \forall t \geq t_0. \tag{7.3}$$

By integration, for all $t \geq t_0$, it yields

$$\frac{1}{2}\|x(t) - x_*\|^2 + \int_{t_0}^{t} (f(x(s)) - f_*)ds \leq \frac{1}{2}\|x(t_0) - x_*\|^2 = \frac{1}{2}\|x^0 - x_*\|^2.$$

Consequently, $\int_{t_0}^{+\infty} (f(x(t)) - f_*)dt < +\infty$, and statement (a) is proved.

From $\int_{t_0}^{+\infty} \frac{1}{t} t(f(x(t)) - f_*)dt = \int_{t_0}^{+\infty} (f(x(t)) - f_*)dt < +\infty$, we immediately deduce that $\liminf_{t\to+\infty} t(f(x(t)) - f_*) = 0$. Indeed, assuming that $\liminf_{t\to+\infty} t(f(x(t)) - f_*) > 0$, there exist $c > 0$ and $t_1 > t_0$ such that $t(f(x(t)) - f_*) > c$ for all $t \geq t_1$. This implies

$$+\infty > \int_{t_0}^{+\infty} (f(x(t)) - f_*)dt \geq \int_{t_1}^{+\infty} \frac{c}{t} = \lim_{t\to+\infty} (\ln(t) - \ln(t_1)) = +\infty,$$

and leads to a contradiction.

For all $t \geq t_0$ it holds

$$\frac{d}{dt}\left(t(f(x(t)) - f_*)\right) = f(x(t)) - f_* + t\nabla f(x(t))^T \dot{x}(t) = f(x(t)) - f_* - t\|\dot{x}(t)\|^2 \leq f(x(t)) - f_*.$$

Since $t \mapsto t(f(x(t)) - f_*)$ is bounded from below and $t \mapsto f(x(t)) - f_*$ is integrable, by Lemma 7.1, we obtain that $\lim_{t\to+\infty} t(f(x(t)) - f_*)$ exists and is finite. This means that $\lim_{t\to+\infty} t(f(x(t)) - f_*) = 0$, which completes the proof of (b).

Now, we will turn our attention to statement (c). According to (7.3), we have

$$\lim_{t\to+\infty} \|x(t) - x_*\|^2 := \ell_{x^*} \quad \text{exists and is finite.}$$

Since $x(\cdot)$ is bounded, it has at least one accumulation point. We will prove that it has exactly one accumulation point, which implies that the trajectory converges.

Indeed, assume that $x(\cdot)$ has two accumulation points, namely, x' and x''. Then there exist subsequences $x(t_l)_{l\geq 0}$ and $x(t_j)_{j\geq 0}$ such that

$$x(t_l) \to x' \ (l \to +\infty) \quad \text{and} \quad x(t_j) \to x'' \ (j \to +\infty).$$

The continuity of f and (b) yield

$$\begin{aligned} f(x(t_l)) \to f(x') \ (l \to +\infty) &\ \Rightarrow\ f(x') = f_* \ \Rightarrow\ x' \in \operatorname{argmin} f \\ f(x(t_j)) \to f(x'') \ (j \to +\infty) &\ \Rightarrow\ f(x'') = f_* \ \Rightarrow\ x'' \in \operatorname{argmin} f. \end{aligned}$$

For every $t \geq t_0$ we have

$$2x(t)^T(x' - x'') = \|x(t) - x''\|^2 - \|x(t) - x'\|^2 - \|x''\|^2 + \|x'\|^2,$$

then

$$2x(t_l)^T(x' - x'') = \|x(t_l) - x''\|^2 - \|x(t_l) - x'\|^2 - \|x''\|^2 + \|x'\|^2 \ \ \forall l \geq 0$$
$$2x(t_j)^T(x' - x'') = \|x(t_j) - x''\|^2 - \|x(t_j) - x'\|^2 - \|x''\|^2 + \|x'\|^2 \ \ \forall j \geq 0.$$

We let $l \to +\infty$ and $j \to +\infty$, respectively, and so

$$2(x')^T(x' - x'') = l_{x''} - l_{x'} - \|x''\|^2 - \|x'\|^2$$
$$2(x'')^T(x' - x'') = l_{x''} - l_{x'} - \|x''\|^2 - \|x'\|^2,$$

which implies

$$(x')^T(x' - x'') = (x'')^T(x' - x'') \Leftrightarrow \ (x' - x'')^T(x' - x'') = 0 \Leftrightarrow \ \|x' - x''\|^2 = 0 \ \Leftrightarrow \ x' = x''.$$

Thus, the trajectory $x(t)$ is convergent to an element in $\operatorname{argmin} f$ as $t \to +\infty$.

The finiteness of the integral $\int_{t_0}^{+\infty} \frac{1}{t} t^2 \|\nabla f(x(t))\|^2 dt = \int_{t_0}^{+\infty} t\|\nabla f(x(t))\|^2 dt < +\infty$ yields that $\liminf_{t\to+\infty} t^2\|\nabla f(x(t))\|^2 = 0$. Since $t \mapsto x(t)$ is locally absolutely continuous and ∇f is Lipschitz continuous (on bounded sets), we obtain that $t \mapsto \nabla f(x(t))$ is locally absolutely continuous. Therefore, for almost every $t \geq t_0$, it holds

$$\begin{aligned}\frac{d}{dt}\left(\frac{1}{2}t^2\|\nabla f(x(t))\|^2\right) &= t\|\nabla f(x(t))\|^2 + t^2\nabla f(x(t))^T\left(\frac{d}{dt}\nabla f(x(t))\right) \\ &= t\|\nabla f(x(t))\|^2 - t^2\dot{x}(t)^T\left(\frac{d}{dt}\nabla f(x(t))\right).\end{aligned}$$

Since ∇f is monotone, for all $t_0 \leq t < s$, it holds

$$(x(s) - x(t))^T(\nabla f(x(s)) - \nabla f(x(t))) \geq 0,$$

therefore,

$$\left(\frac{x(s) - x(t)}{s - t}\right)^T\left(\frac{\nabla f(x(s)) - \nabla f(x(t))}{s - t}\right) \geq 0.$$

We let $s \downarrow t$ and obtain from here that $\dot{x}(t)^T\left(\frac{d}{dt}\nabla f(x(t))\right) \geq 0$ for almost every $t \geq t_0$. This means that, for almost every $t \geq t_0$,

$$\frac{d}{dt}\left(\frac{1}{2}t^2\|\nabla f(x(t))\|^2\right) \leq t\|\nabla f(x(t))\|^2.$$

Applying again Lemma 7.1 and using (a), we obtain that $\lim_{t\to+\infty} t^2\|\nabla f(x(t))\|^2$ exists, therefore, $\lim_{t\to+\infty} t^2\|\nabla f(x(t))\|^2 = 0$. This concludes the proof of statement (d). ■

7.2 The gradient algorithm for convex optimization problems

The gradient algorithm that we will introduce for solving (7.1) can be seen as an explicit time discretization of the gradient flow system

$$\begin{cases} \dot{x}(t) = -\nabla f(x(t)) \\ x(t_0) = x^0. \end{cases}$$

Indeed, for all $k \geq 0$, we have

$$\frac{x(t_{k+1}) - x(t_k)}{\gamma_k} = -\nabla f(x(t_k)).$$

Setting $x(t_0) := x^0$, $x(t_k) := x^k$ and $\gamma_k := \gamma$ for all $k \geq 0$, we get

$$x^{k+1} - x^k = -\gamma \nabla f(x^k) \;\Leftrightarrow\; x^{k+1} = x^k - \gamma \nabla f(x^k),$$

which is the update rule of the gradient method with constant step size.

Lemma 7.3 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex and differentiable function with a $L_{\nabla f}$-Lipschitz continuous gradient.*

(a) (Descent Lemma) *For all $x, y \in \mathbb{R}^n$, it holds*

$$f(y) \leq f(x) + \nabla f(x)^T (y - x) + \frac{L_{\nabla f}}{2} \|y - x\|^2 .$$

(b) (Baillon-Haddad Theorem) *For all $x, y \in \mathbb{R}^n$, it holds*

$$L_{\nabla f} (\nabla f(y) - \nabla f(x))^T (y - x) \geq \|\nabla f(y) - \nabla f(x)\|^2.$$

Proof. (a) Let $x, y \in \mathbb{R}^n$ and define $\phi : [0,1] \to \mathbb{R}^n, \phi(t) = f(x + t(y - x))$. For every $t \in [0,1]$ it holds $\phi'(t) = \nabla f(x + t(y - x))^T (y - x)$. From the fundamental theorem of differentiation and integration we get

$$\phi(1) - \phi(0) = \int_0^1 \phi'(t) dt \;\Leftrightarrow\; f(y) - f(x) = \int_0^1 \nabla f(x + t(y - x))^T (y - x) dt,$$

which gives

$$\begin{aligned} f(y) - f(x) - \nabla f(x)^T (y - x) &= \int_0^1 \Big(\nabla f(x + t(y - x)) - \nabla f(x) \Big)^T (y - x) \, dt \\ &\leq \int_0^1 \left| \Big(\nabla f(x + t(y - x)) - \nabla f(x) \Big)^T (y - x) \right| dt \\ &\leq \int_0^1 \|\nabla f(x + t(y - x)) - \nabla f(x)\| \, \|y - x\| \; dt \\ &\leq \int_0^1 L_{\nabla f} t \, \|y - x\|^2 \; dt \\ &= L_{\nabla f} \, \|y - x\|^2 \int_0^1 t \; dt = \frac{L_{\nabla f}}{2} \|y - x\|^2 . \end{aligned}$$

(b) Let $x, y \in \mathbb{R}^n$. For all $z \in \mathbb{R}^n$, by applying the gradient inequality and the Descent Lemma, we have

$$\begin{aligned} & f(y) - f(x) - \nabla f(x)^T (y - x) \\ \geq\ & f(y) - f(z) + \nabla f(x)^T (z - x) - \nabla f(x)^T (y - x) \\ \geq\ & -\nabla f(y)^T (z - y) - \frac{L_{\nabla f}}{2} \|y - z\|^2 + \nabla f(x)^T (z - y) \\ =\ & (\nabla f(x) - \nabla f(y))^T (z - y) - \frac{L_{\nabla f}}{2} \|y - z\|^2 . \end{aligned}$$

This means that

$$\begin{aligned} f(y) - f(x) - \nabla f(x)^T (y - x) &\geq \sup_{z \in \mathbb{R}^n} \left((\nabla f(x) - \nabla f(y))^T (z - y) - \frac{L_{\nabla f}}{2} \|y - z\|^2 \right) \\ &\geq \sup_{u \in \mathbb{R}^n} \left((\nabla f(x) - \nabla f(y))^T u - \frac{L_{\nabla f}}{2} \|u\|^2 \right) \\ &= \frac{1}{2 L_{\nabla f}} \|\nabla f(x) - \nabla f(y)\|^2 . \end{aligned}$$

By interchanging the roles of x and y and summing the resulting inequalities, we obtain

$$L_{\nabla f} (\nabla f(y) - \nabla f(x))^T (y - x) \geq \|\nabla f(y) - \nabla f(x)\|^2 .$$

■

Remark 7.4 Under the assumptions of Lemma 7.3 we have

$$f(x) + \nabla f(x)^T (y - x) \leq f(y) \leq f(x) + \nabla f(x)^T (y - x) + \frac{L_{\nabla f}}{2} \|y - x\|^2 \quad \forall x, y \in \mathbb{R}^n .$$

This means that the graph of f is "squeezed" between an affine function and a quadratic function.

The following lemma is useful in analyzing the convergence of algorithms. It can be seen as the discrete counterpart of Lemma 7.1.

Lemma 7.5 (Robbins-Monro Lemma) *Let $(a_k)_{k \geq 0}$, $(b_k)_{k \geq 0}$ and $(d_k)_{k \geq 0}$ be sequences of real numbers such that $(a_k)_{k \geq 0}$ is bounded from below, $(b_k)_{k \geq 0}$ is nonnegative and $(d_k)_{k \geq 0} \in \ell_1$. If*

$$a_{k+1} \leq a_k - b_k + d_k \quad \forall k \geq 0,$$

then $\lim_{k \to +\infty} a_k$ exists and is finite, and $(b_k)_{k \geq 0} \in \ell_1$.

Proof. Using that $\sum_{k=0}^{+\infty} \max\{d_k, 0\} \leq \sum_{k=0}^{+\infty} |d_k| < +\infty$ and $\sum_{k=0}^{+\infty} \max\{-d_k, 0\} \leq \sum_{k=0}^{+\infty} |d_k| < +\infty$, we have

$$\lim_{k \to +\infty} \sum_{l=0}^{k} d_l = \sum_{k=0}^{+\infty} \max\{d_k, 0\} + \sum_{k=0}^{+\infty} \max\{-d_k, 0\} \text{ exists and is finite.}$$

For all $k \geq 1$, we define

$$c_k := a_k - \sum_{l=0}^{k-1} d_l.$$

Then, for all $k \geq 1$, it holds

$$\inf_{l\geq 0} a_l - \sum_{l=0}^{k} d_l \leq c_{k+1} = a_{k+1} - \sum_{l=0}^{k} d_l \leq a_k - b_k + d_k - \sum_{l=0}^{k} d_l = a_k - b_k - \sum_{l=0}^{k-1} d_l = c_k - b_k \leq c_k.$$

Therefore, $\lim_{k\to+\infty} c_k$ exists and is finite, which implies that $\lim_{k\to+\infty} a_k$ exists and is finite.

On the other hand, for all $k \geq 0$, it holds

$$\sum_{l=0}^{k} b_l \leq a_0 - a_{k+1} + \sum_{l=0}^{k} d_l \leq a_0 - \inf_{l\geq 0} a_l + \sum_{l=0}^{+\infty} d_l,$$

which yields $(b_k)_{k\geq 0} \in \ell_1$. ■

Theorem 7.6 (convergence of the gradient method for convex optimization problems) *Let $x^0 \in \mathbb{R}^n$ be an arbitrary starting point and $\gamma \in \left(0, \frac{1}{L_{\nabla f}}\right]$. For the sequence $(x^k)_{k\geq 0}$ generated by the gradient method with constant step size γ,*

$$x^{k+1} := x^k - \gamma \nabla f(x^k) \quad \forall k \geq 0,$$

the following statements are true:

(a) it holds

$$\left(k\|x^{k+1} - x^k\|^2\right)_{k\geq 0} \in \ell_1, \left(k\|\nabla f(x^k)\|^2\right)_{k\geq 0} \in \ell_1 \quad \text{and} \quad \left(f(x^k) - f_*\right)_{k\geq 0} \in \ell_1;$$

(b) it holds $f\left(x^k\right) - f_* = o\left(\frac{1}{k}\right)$ as $k \to +\infty$;

(c) $(x^k)_{k\geq 0}$ converges to an element in $\operatorname{argmin} f$ as $k \to +\infty$;

(d) it holds $\|\nabla f(x^k)\| = o\left(\frac{1}{k}\right)$ as $t \to +\infty$;

Proof. For $x^* \in \operatorname{argmin} f$, we define the discrete energy function

$$\mathcal{E}_k := k\left(f\left(x_k\right) - f_*\right) + \frac{1}{2\gamma}\left\|x_k - x^*\right\|^2 \quad \forall k \geq 0.$$

For all $x \in \mathbb{R}^n$ and all $k \geq 0$, we have by the gradient inequality

$$f(x) - f(x^k) \geq \nabla f(x^k)^T (x - x^k),$$

while Lemma 7.3 yields

$$\begin{aligned} f(x^{k+1}) &\leq f(x^k) + \nabla f(x^k)^T(x^{k+1} - x^k) + \frac{L_{\nabla f}}{2}\left\|x^{k+1} - x^k\right\|^2 \\ \Leftrightarrow\ f(x^k) - f(x^{k+1}) &\geq \nabla f(x^k)^T(x^k - x^{k+1}) - \frac{L_{\nabla f}}{2}\left\|x^{k+1} - x^k\right\|^2. \end{aligned}$$

Adding the two inequalities, for all $x \in \mathbb{R}^n$ and all $k \geq 0$, we get

$$f(x) - f(x^{k+1}) \geq \nabla f(x^k)^T(x - x^{k+1}) - \frac{L_{\nabla f}}{2}\left\|x^{k+1} - x^k\right\|^2$$

and further, after substituting $\frac{1}{\gamma}(x^k - x^{k+1})$ for $\nabla f(x^k)$,

$$f(x) - f(x^{k+1}) \geq \frac{1}{\gamma}(x^k - x^{k+1})^T(x - x^{k+1}) - \frac{L_{\nabla f}}{2}\left\|x^{k+1} - x^k\right\|^2.$$

In other words, for all $x \in \mathbb{R}^n$ and all $k \geq 0$, it holds

$$f(x) - f(x^{k+1}) \geq \frac{1}{2\gamma}\left(\left\|x^{k+1} - x^k\right\|^2 + \left\|x - x^{k+1}\right\|^2 - \left\|x - x^k\right\|^2\right) - \frac{L_{\nabla f}}{2}\left\|x^{k+1} - x^k\right\|^2,$$

which is equivalent to

$$f(x^{k+1}) - f(x) + \frac{1}{2\gamma}(\|x^{k+1} - x\|^2 - \|x^k - x\|^2) + \frac{1 - \gamma L_{\nabla f}}{2\gamma}\|x^{k+1} - x^k\|^2 \leq 0.$$

From here we deduce, for all $k \geq 0$, that (by setting $x := x^k$)

$$f(x^{k+1}) - f(x^k) \leq f(x^{k+1}) - f(x^k) + \frac{2 - \gamma L_{\nabla f}}{2\gamma}\|x^{k+1} - x^k\|^2 \leq 0, \tag{7.4}$$

and (by setting $x := x^*$)

$$f(x^{k+1}) - f_* + \frac{1}{2\gamma}(\|x^{k+1} - x^*\|^2 - \|x^k - x^*\|^2) + \frac{1 - \gamma L_{\nabla f}}{2\gamma}\|x^{k+1} - x^k\|^2 \leq 0. \tag{7.5}$$

This yields, for all $k \geq 0$,

$$\begin{aligned} \mathcal{E}_{k+1} - \mathcal{E}_k &= k\big(f(x^{k+1}) - f(x^k)\big) + f(x^{k+1}) - f_* + \frac{1}{2\gamma}(\|x^{k+1} - x^*\|^2 - \|x^k - x^*\|^2) \\ &\leq\ -k\frac{2 - \gamma L_{\nabla f}}{2\gamma}\|x^{k+1} - x^k\|^2 - \frac{1 - \gamma L_{\nabla f}}{2\gamma}\|x^{k+1} - x^k\|^2, \end{aligned}$$

and so, using that $\gamma L_{\nabla f} \leq 1$,

$$\begin{aligned} \mathcal{E}_{k+1} - \mathcal{E}_k + k\frac{2 - \gamma L_{\nabla f}}{2\gamma}\|x^{k+1} - x^k\|^2 &\leq \mathcal{E}_{k+1} - \mathcal{E}_k + \frac{1}{2\gamma}\left(k(2 - \gamma L_{\nabla f}) + 1 - \gamma L_{\nabla f}\right)\|x^{k+1} - x^k\|^2 \\ &\leq 0. \end{aligned} \tag{7.6}$$

By using telescoping arguments, we obtain, for all $K \geq 1$,

$$K\left(f\left(x^K\right)-f_*\right)+\frac{1}{2\gamma}\left\|x^K-x^*\right\|^2+\frac{2-\gamma L}{2\gamma}\sum_{k=0}^{K-1}k\|x^{k+1}-x^k\|^2\leq\frac{1}{2\gamma}\left\|x^0-x^*\right\|^2.$$

This yields that $(x^k)_{k\geq 0}$ is bounded, $(k\|x_{k+1}-x_k\|^2)_{k\geq 0}\in\ell_1$, $(k\|\nabla f(x_k)\|^2)_{k\geq 0}\in\ell_1$, and

$$0\leq f(x^k)-f_*\leq\frac{\left\|x^0-x^*\right\|^2}{2\gamma k}\quad\forall k\geq 1. \tag{7.7}$$

On the other hand, from (7.5), we obtain that, for all $k\geq 0$,

$$\sum_{l=0}^{k}(f(x^{k+1})-f_*)\leq\frac{1}{2\gamma}\left\|x^0-x^*\right\|^2,$$

thus $\left(f(x^k)-f_*\right)_{k\geq 0}\in\ell_1$. This proves statement (a).

Since $\sum_{k=1}^{+\infty}\frac{1}{k}k(f(x_k)-f_*)=\sum_{k=1}^{+\infty}(f(x_k)-f_*)<+\infty$, it holds $\liminf_{k\to+\infty}k(f(x_k)-f_*)=0$. In addition, from (7.4), for all $k\geq 0$, we have

$$(k+1)(f(x^{k+1})-f_*)\leq k(f(x^k)-f_*)+(f(x^k)-f_*).$$

Lemma 7.5, yields that $\lim_{k\to+\infty}k(f(x^k)-f_*)$ exists and it is finite, thus $\lim_{k\to+\infty}k(f(x^k)-f_*)=0$. This completes the proof of statement (b).

Making again use of (7.5), we have that, for all $k\geq 0$,

$$\|x^{k+1}-x^*\|^2\leq\|x^k-x^*\|^2-2\gamma(f(x^{k+1})-f_*)+\frac{\gamma L_{\nabla f}-1}{2\gamma}\|x^{k+1}-x^k\|^2.$$

Lemma 7.5 yields that $\lim_{k\to+\infty}\|x^k-x^*\|^2$ exists and it is finite – let

$$\ell_{x^*}:=\lim_{k\to+\infty}\left\|x^k-x^*\right\|^2.$$

This implies that $(x^k)_{k\geq 0}$ is bounded, and therefore it has at least one accumulation point. We will prove that it has exactly one accumulation point, which implies that the whole sequence converges.

Indeed, assume that $(x^k)_{k\geq 0}$ has two accumulation points x' and x''. Then there exist subsequences $(x^{k_l})_{l\geq 0}$ and $(x^{k_j})_{j\geq 0}$ such that

$$x^{k_l}\to x'\ (l\to+\infty)\quad\text{and}\quad x^{k_j}\to x''\ (j\to+\infty).$$

By continuity of f and by part (b), we have that

$$\begin{aligned}&f(x^{k_l})\to f(x')\ (l\to+\infty)\ \Rightarrow\ f(x')=f^*\ \Rightarrow\ x'\in\operatorname{argmin}f\\&f(x^{k_j})\to f(x'')\ (j\to+\infty)\ \Rightarrow\ f(x'')=f^*\ \Rightarrow\ x''\in\operatorname{argmin}f.\end{aligned}$$

For every $k \geq 0$ we have

$$2(x^k)^T(x' - x'') = \left\|x^k - x''\right\|^2 - \left\|x^k - x'\right\|^2 - \|x''\|^2 + \|x'\|^2,$$

then

$$2(x^{k_l})^T(x' - x'') = \left\|x^{k_l} - x''\right\|^2 - \left\|x^{k_l} - x'\right\|^2 - \|x''\|^2 + \|x'\|^2 \ \ \forall l \geq 0$$
$$2(x^{k_j})^T(x' - x'') = \left\|x^{k_j} - x''\right\|^2 - \left\|x^{k_j} - x'\right\|^2 - \|x''\|^2 + \|x'\|^2 \ \ \forall j \geq 0.$$

We let $l \to +\infty$ and $j \to +\infty$, respectively, and so

$$2(x')^T(x' - x'') = l_{x''} - l_{x'} - \|x''\|^2 - \|x'\|^2$$
$$2(x'')^T(x' - x'') = l_{x''} - l_{x'} - \|x''\|^2 - \|x'\|^2,$$

which implies

$$(x')^T(x' - x'') = (x'')^T(x' - x'') \Leftrightarrow \ (x' - x'')^T(x' - x'') = 0 \Leftrightarrow \ \|x' - x''\|^2 = 0 \ \Leftrightarrow \ x' = x''.$$

Thus, the whole sequence $(x^k)_{k\geq 0}$ is convergent to an element in $\operatorname{argmin} f$.

According to (a), $\sum_{k=1}^{+\infty} \frac{1}{k} k^2 \|\nabla f(x^k)\|^2 = \sum_{k=1}^{+\infty} k\|\nabla f(x^k)\|^2 < +\infty$. From here, we conclude that $\liminf_{k\to+\infty} k\|\nabla f(x^k)\| = 0$. By using the Baillon-Haddad Theorem, for all $k \geq 0$, we have

$$\begin{aligned}\|\nabla f(x^{k+1}) - \nabla f(x^k)\|^2 &\leq L_{\nabla f}\left(\nabla f(x^{k+1}) - \nabla f(x^k)\right)^T (x^{k+1} - x^k) \\ &= -\gamma L_{\nabla f}\left(\nabla f(x^{k+1}) - \nabla f(x^k)\right)^T \nabla f(x^k) \\ &\leq -2\left(\nabla f(x^{k+1}) - \nabla f(x^k)\right)^T \nabla f(x^k),\end{aligned}$$

from where we deduce that

$$\|\nabla f(x^{k+1})\|^2 \leq \|\nabla f(x^k)\|^2.$$

This allows us to conclude that, for all $k \geq 0$,

$$\begin{aligned}(k+1)^2\|\nabla f(x^{k+1})\|^2 = (k^2 + 2k + 1)\|\nabla f(x^{k+1})\|^2 &\leq k^2\|\nabla f(x^k)\|^2 + (2k+1)\|\nabla f(x^{k+1})\|^2 \\ &\leq 2(k+1)\|\nabla f(x^{k+1})\|^2.\end{aligned}$$

By Lemma 7.5, $\lim_{k\to+\infty} k^2\|\nabla f(x^k)\|^2$ exists and it is finite, thus $\lim_{k\to+\infty} k\|\nabla f(x^k)\| = 0$. This completes the proof of statement (d). ■

Remark 7.7 By slightly modifying the arguments in the previous proof, one can show that the conclusions of Theorem 7.6 remain valid for all $\gamma \in \left(0, \frac{2}{L_{\nabla f}}\right)$.

Example 7.8 We can use the gradient method to solve a linear system $Ax = b$ for $A \in \mathbb{R}^{m\times n}$, $A \neq 0$, and $b \in \mathbb{R}^m$, by reformulating it as

$$\min_{x\in\mathbb{R}^n} \|Ax - b\|^2.$$

Define $f : \mathbb{R}^n \to \mathbb{R}$, $f(x) = \|Ax - b\|^2$. Then f is convex and differentiable and its gradient $x \mapsto \nabla f(x) = 2A^T(Ax - b)$ is Lipschitz continuous with constant $L_{\nabla f} := 2\left\|A^T A\right\| = 2\|A\|^2$. Indeed, we have

$$\|\nabla f(x) - \nabla f(y)\| = 2\left\|A^T A(x - y)\right\| \leq 2\left\|A^T A\right\| \|x - y\| \quad \forall x, y \in \mathbb{R}^n.$$

Therefore, for $\gamma \in \left(0, \frac{1}{2\|A\|^2}\right]$, we can solve $Ax = b$ with the iteration

$$x^{k+1} := x^k - 2\gamma A^T(Ax^k - b) \quad \forall k \geq 0.$$

7.3 The fast gradient method for convex optimization problems

In 1983, Nesterov introduced in [8] the following so-called accelerated/fast gradient method for solving (7.1): for $x^0 = x^1 \in \mathbb{R}^n$, $\gamma \in \left(0, \frac{1}{L_{\nabla f}}\right]$,

$$t_1 = 1 \quad \text{and} \quad t_{k+1} := \frac{1 + \sqrt{4t_k^2 + 1}}{2} \quad \forall k \geq 1$$

(notice that $t_{k+1}^2 - t_{k+1} - t_k^2 = 0$), let

$$(\forall k \geq 1) \quad \begin{cases} y^k := x^k + \dfrac{t_k - 1}{t_{k+1}}(x^k - x^{k-1}) \\ x^{k+1} := y^k - \gamma \nabla f(y^k). \end{cases} \tag{7.8}$$

The sequence $(y^k)_{k \geq 1}$ is called the momentum sequence. For every $k \geq 2$ it holds

$$0 \leq f(x^k) - f^* \leq \frac{2}{\gamma(k+1)^2} \operatorname{dist}^2_{\operatorname{argmin} f}(x^0),$$

which means that $(f(x^k))_{k \geq 0}$ converges to f^* with a convergence rate of $\mathcal{O}(\frac{1}{k^2})$ as $k \to +\infty$, which is faster than for the gradient method. However, it is not known whether the sequence of iterates $(x^k)_{k \geq 0}$ converges.

In 2015, Chambolle-Dossal introduced in [3] the following modified version of Nesterov's accelerated/fast gradient method: for $x^0 = x^1 \in \mathbb{R}^n$, $\gamma \in \left(0, \frac{1}{L_{\nabla f}}\right]$,

$$\alpha \geq 3 \quad \text{and} \quad t_k := \frac{k + \alpha - 2}{\alpha - 1} \quad \forall k \geq 1$$

(notice that $t_{k+1}^2 - t_{k+1} - t_k^2 \leq 0$), let

$$(\forall k \geq 1) \quad \begin{cases} y^k := x^k + \dfrac{t_k - 1}{t_{k+1}}(x^k - x^{k-1}) \\ x^{k+1} := y^k - \gamma \nabla f(y^k). \end{cases} \tag{7.9}$$

The following holds:

(i) $(f(x^k))_{k\geq 0}$ converges to f^* with a convergence rate of $\mathcal{O}(\frac{1}{k^2})$ as $k \to +\infty$;

(ii) if $\alpha > 3$, then $(f(x^k))_{k\geq 0}$ converges to f^* with a convergence rate of $o(\frac{1}{k^2})$, and the sequence of $(x^k)_{k\geq 0}$ converges to an element of argmin f as $k \to +\infty$.

The convergence of the iterates to a global minimizer of f, for the fast gradient method with Nesterov momentum and with Chambolle-Dossal momentum when $\alpha = 3$, was shown in [2] in October 2025.

The continuous counterpart of (7.9) has the following formulation

$$\begin{cases} \ddot{x}(t) + \dfrac{\alpha}{t}\dot{x}(t) + \nabla f(x(t)) = 0 \\ x(t_0) = x^0, \dot{x}(t_0) = \dot{x}^0, \end{cases} \tag{7.10}$$

where $t \geq t_0 > 0$ and $\alpha \geq 3$. It was proposed by Su-Boyd-Candés in 2015 for $\alpha = 3$. The following holds, as in discrete time,

(i) $f(x(t)) \to f^*$ with a convergence rate of $\mathcal{O}(\frac{1}{t^2})$ as $t \to +\infty$;

(ii) if $\alpha > 3$, then $f(x(t)) \to f^*$ with a convegrence rate of $o(\frac{1}{t^2})$, and $x(t)$ converges to an element of argmin f as $t \to +\infty$.

The convergence of the trajectory in case $\alpha = 3$ was shown in [7] in October 2025.

Remark 7.9 For $\alpha = 3$, the fast convergence rate for the objective function values can be obtained by considering, for $x^* \in \operatorname{argmin} f$, the following energy function

$$\mathcal{E} : [t_0, +\infty) \to \mathbb{R}, \quad \mathcal{E}(t) = t^2(f(x(t)) - f_*) + \frac{1}{2}\|t\dot{x}(t) + 2(x(t) - x_*)\|^2.$$

For all $t \geq t_0$, it holds

$$\begin{aligned} \frac{d}{dt}\mathcal{E}(t) &= 2t(f(x(t)) - f_*) + t^2\nabla f(x(t))^T\dot{x}(t) + (t\dot{x}(t) + 2(x(t) - x^*))^T(t\ddot{x}(t) + 3\dot{x}(t)) \\ &= 2t(f(x(t)) - f_*) + t^2\nabla f(x(t))^T\dot{x}(t) + (t\dot{x}(t) + 2(x(t) - x^*))^T(-t\nabla f(x(t))) \\ &= 2t(f(x(t)) - f_* - (x(t) - x^*)\nabla f(x(t))) \leq 0. \end{aligned}$$

Hence, for all $t \geq t_0$,

$$0 \leq f(x(t)) - f_* \leq \frac{\mathcal{E}(t)}{t^2} \leq \frac{\mathcal{E}(t_0)}{t^2}.$$

7.4 The minimization of a strongly convex function

In the following, we will discuss the convergence properties of the gradient method when minimizing a strongly convex function.

Definition 7.10 (strongly convex function) A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be strongly convex (with modulus $\mu > 0$) if

$$f(\lambda x+(1-\lambda)y)+\mu\lambda(1-\lambda)\|x-y\|^2 \leq \lambda f(x)+(1-\lambda)f(y) \text{ for every } x, y \in \mathbb{R}^n \text{ and every } \lambda \in [0,1].$$

The function f is strongly convex with modulus $\mu > 0$ if and only if

$$g : \mathbb{R}^n \to \mathbb{R},\ g(x) = f(x) - \mu\|x\|^2,$$

is convex. Every strongly convex function is strictly convex and it has a unique global minimum.

The following theorem characterizes the convergence properties of the gradient method with fixed step size when applied to the minimization of a strongly convex function. Notice that if $f : \mathbb{R}^n \to \mathbb{R}$ is a strongly convex with modulus $\mu > 0$ and differentiable function with a $L_{\nabla f}$-Lipschitz continuous gradient, then, according to Lemma 7.3, it holds

$$\mu\|x-y\|^2 + \nabla f(x)^T(y-x) + f(x) \leq f(y) \leq f(x) + \nabla f(x)^T(y-x) + \frac{L_{\nabla f}}{2}\|x-y\|^2 \ \forall x, y \in \mathbb{R}^n,$$

therefore $\mu \leq \frac{L_{\nabla f}}{2}$.

Theorem 7.11 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a strongly convex with modulus $\mu > 0$ and differentiable function with a $L_{\nabla f}$-Lipschitz continuous gradient, $x^0 \in \mathbb{R}^n$, $\gamma \in \left(0, \frac{4\mu}{L^2_{\nabla f}}\right)$ and $c := 1-4\gamma\mu+\gamma^2 L^2_{\nabla f} \in [0,1)$. For the sequence $(x^k)_{k\geq 0}$ generated by the gradient method with constant step size γ,*

$$x^{k+1} := x^k - \gamma\nabla f(x^k) \quad \forall k \geq 0,$$

the following statements are true:

(a) for every $k \geq 0$, $\left\|x^{k+1} - x^*\right\| \leq \sqrt{c}\left\|x^k - x^*\right\|$, where x^* is the unique global minimum of f;

(b) $(x^k)_{k\geq 0}$ converges to x^* with a convergence rate of $\mathcal{O}(\sqrt{c}^k)$ as $k \to +\infty$;

(c) if $\gamma \in \left(0, \frac{2\mu}{L^2_{\nabla f}}\right]$, then $(f(x^k))_{k\geq 0}$ converges to f^* with a convergence rate of $\mathcal{O}(c^k)$ as $k \to +\infty$.

Proof. Since $\mu \leq \frac{L_{\nabla f}}{2}$, it holds $c \geq 1-2\gamma L_{\nabla f}+\gamma^2 L^2_{\nabla f} = (1-\gamma L_{\nabla f})^2 \geq 0$, whereas the condition $\gamma \in \left(0, \frac{4\mu}{L^2_{\nabla f}}\right)$ guarantees that $c < 1$.

(a) Let $T : \mathbb{R}^n \to \mathbb{R}^n$, $T(x) = x - \gamma\nabla f(x)$. We will prove that T is a contraction. The function f is strongly convex with modulus μ, therefore $f - \mu\|\cdot\|^2$ is convex and we get from the gradient inequality that for every $x, y \in \mathbb{R}^n$

$$\begin{aligned} & \left(\nabla f(y) - 2\mu y - \nabla f(x) + 2\mu x\right)^T (y-x) \geq 0 \\ \Leftrightarrow\ & \left(\nabla f(y) - \nabla f(x)\right)^T (y-x) \geq 2\mu(y-x)^T(y-x) = 2\mu\|y-x\|^2, \end{aligned}$$

Then we have for every $x, y \in \mathbb{R}^n$

$$\begin{aligned}\|T(x) - T(y)\|^2 &= \|x - y - \gamma(\nabla f(x) - \nabla f(y))\|^2 \\ &= \|x-y\|^2 - 2\gamma\left(\nabla f(x) - \nabla f(y)\right)^T (x-y) + \gamma^2 \|\nabla f(x) - \nabla f(y)\|^2 .\end{aligned}$$

Since $\left(\nabla f(x) - \nabla f(y)\right)^T (x-y) \geq 2\mu \|y-x\|^2$ and $\|\nabla f(x) - \nabla f(y)\|^2 \leq L_{\nabla f}^2 \|x-y\|^2$, it yields

$$\|T(x) - T(y)\| \leq \|x-y\| \sqrt{1 - 4\gamma\mu + \gamma^2 L_{\nabla f}^2} = \sqrt{c}\,\|x-y\| \quad \forall x, y \in \mathbb{R}^n,$$

which proves that T is a contraction.

According to the Banach-Picard Theorem, the sequence generated by $x^{k+1} := T(x^k)$ for every $k \geq 0$ converges to the unique fixed point of T, which we denote by x^*. In other words,

$$x^* = T(x^*) = x^* - \gamma \nabla f(x^*) \;\Leftrightarrow\; \nabla f(x^*) = 0,$$

which is equivalent to x^* being global minimum of f. Furthermore, we get for every$k \geq 0$

$$\left\|x^{k+1} - x^*\right\| = \left\|T(x^k) - T(x^*)\right\| \leq \sqrt{c}\left\|x^k - x^*\right\|.$$

(b) Follows directly from (a) by applying the inequality k times.

(c) Since $\mu \leq \frac{L_{\nabla f}}{2}$, it holds $\gamma \in \left(0, \frac{1}{L_{\nabla f}}\right)$. According to inequality (7.5), we get

$$0 \leq f(x^{k+1}) - f^* = f(x^{k+1}) - f(x^*) \leq \frac{1}{2\gamma}\left\|x^k - x^*\right\|^2 \leq \frac{1}{2\gamma} c^k \left\|x^0 - x^*\right\|^2 = \frac{\|x^0 - x^*\|^2}{2\gamma c} c^{k+1},$$

which proves the claim. ■

8 The Newton method

8.1 Convergence rates

Definition 8.1 Let $(x^k)_{k\geq 0} \subseteq \mathbb{R}^n$ be a given sequence.

(a) The sequence $(x^k)_{k\geq 0}$ is said to converge linearly to $x^* \in \mathbb{R}^n$ if there exist $q \in (0,1)$ and $k_0 \geq 0$ such that

$$\|x^{k+1} - x^*\| \leq q\|x^k - x^*\| \quad \forall k \geq k_0.$$

(b) The sequence $(x^k)_{k\geq 0}$ is said to converge superlinearly to $x^* \in \mathbb{R}^n$ if there exists a sequence $(\varepsilon_k)_{k\geq 0} \downarrow 0$ such that

$$\|x^{k+1} - x^*\| \leq \varepsilon_k \|x^k - x^*\| \quad \forall k \geq 0.$$

(c) If $x^k \to x^* \in \mathbb{R}^n$ as $k \to +\infty$, then $(x^k)_{k\geq 0}$ is said to converge quadratically to $x^* \in \mathbb{R}^n$ if there exists $Q > 0$ such that

$$\|x^{k+1} - x^*\| \leq Q\|x^k - x^*\|^2 \quad \forall k \geq 0.$$

Remark 8.2 Superlinear convergence implies linear convergence and linear convergence implies convergence. However, the last condition alone in Definition 8.1 (c) does not imply convergence, reason why we assume it.

8.2 The Newton algorithm for nonlinear equations

Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be a continuously differentiable mapping. We want to find $x^* \in \mathbb{R}^n$ that solves the nonlinear equation

$$F(x) = 0. \tag{8.1}$$

Assume that x^k is an approximation of x^*. In order to find x^{k+1}, we consider the linearization of F at x^k:

$$F_k : \mathbb{R}^n \to \mathbb{R}^n, \quad F_k(x) = F(x^k) + \nabla F(x^k)(x - x^k), \tag{8.2}$$

where $\nabla F : \mathbb{R}^n \to \mathbb{R}^{n\times n}$ denotes the Jacobian of F. The new iterate x^{k+1} is chosen as a solution of the linear system

$$F_k(x) = 0. \tag{8.3}$$

If $\nabla F(x^k)^{-1}$ exists, then

$$x^{k+1} := x^k - \nabla F(x^k)^{-1} F(x^k).$$

In general, we do not want to calculate the inverse of a matrix explicitly, because it is very costly. We actually need only a solution $d^k \in \mathbb{R}^n$ of the so-called Newton equation

$$\nabla F(x^k) d = -F(x^k) \tag{8.4}$$

and to set afterwards

$$x^{k+1} := x^k + d^k.$$

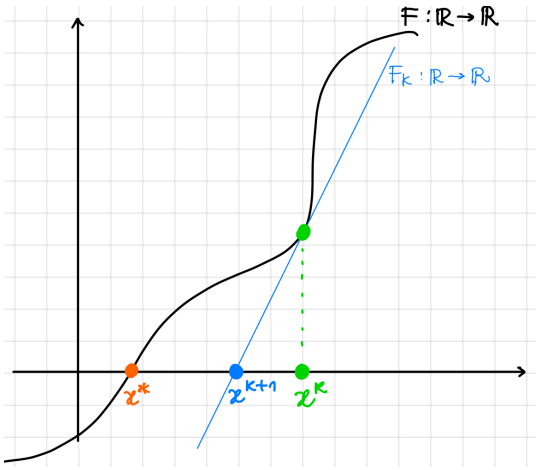

Figure 8.1: One step of the Newton method.

Algorithm 8.3 (Newton algorithm for nonlinear equations)

1: Choose a starting point $x^0 \in \mathbb{R}^n$ and set $k := 0$.

2: If $F(x^k) = 0$: STOP.

3: Find $d^k \in \mathbb{R}^n$ as a solution of the Newton equation

$$\nabla F(x^k)d = -F(x^k).$$

4: Set $x^{k+1} := x^k + d^k$, $k := k+1$ and go to Step 2.

In the following, for a matrix $A \in \mathbb{R}^{n\times n}$ we denote by

$$\|A\| = \max\{\|Ax\| : \|x\| = 1\}$$

the matrix (operator) norm of A induced by the Euclidean norm $\|\cdot\|$.

Lemma 8.4 (Banach Lemma)

(a) Let $M \in \mathbb{R}^{n\times n}$ be a matrix with $\|M\| < 1$. Then $I - M$ is regular and it holds

$$\left\|(I-M)^{-1}\right\| \le \frac{1}{1-\|M\|}.$$

(b) Let $A, B \in \mathbb{R}^{n\times n}$ with $\|I - BA\| < 1$. Then A and B are regular and it holds

$$\left\|A^{-1}\right\| \le \frac{\|B\|}{1-\|I-BA\|} \text{ and } \left\|B^{-1}\right\| \le \frac{\|A\|}{1-\|I-BA\|}$$

The following two lemmas will play an important role in the convergence analysis of the Newton algorithm.

Lemma 8.5 *Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable, $x^* \in \mathbb{R}^n$ and $\nabla F(x^*)$ regular. Then there exist $\varepsilon > 0$ and $c > 0$ such that for every $x \in B(x^*;\varepsilon)$ the matrix $\nabla F(x)$ is regular and $\|\nabla F(x)^{-1}\| \le c$.*

Proof. By the continuity of ∇F, there exists $\varepsilon > 0$ such that for every $x \in B(x^*;\varepsilon)$ it holds

$$\|\nabla F(x) - \nabla F(x^*)\| \le \frac{1}{2\left\|\nabla F(x^*)^{-1}\right\|}.$$

Then we have for every $x \in B(x^*;\varepsilon)$

$$\begin{aligned}\left\|I - \nabla F(x^*)^{-1}\nabla F(x)\right\| &= \left\|\nabla F(x^*)^{-1}(\nabla F(x^*) - \nabla F(x))\right\| \\ &\le \left\|\nabla F(x^*)^{-1}\right\| \|\nabla F(x^*) - \nabla F(x)\| \le \frac{1}{2} < 1.\end{aligned}$$

By Lemma 8.4 (b) we get that for every $x \in B(x^*;\varepsilon)$ the matrix $\nabla F(x)$ is regular and

$$\left\|\nabla F(x)^{-1}\right\| \le \frac{\left\|\nabla F(x^*)^{-1}\right\|}{1-\|I-\nabla F(x^*)^{-1}\nabla F(x)\|} \le 2\left\|\nabla F(x^*)^{-1}\right\| =: c.$$

■

Lemma 8.6 *Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be an operator and $(x^k)_{k\geq 0} \to x^* \in \mathbb{R}^n$ as $k \to +\infty$.*

(a) If F is continuously differentiable, then there exists $(\varepsilon_k)_{k\geq 0} \downarrow 0$ such that

$$\left\|F(x^k) - F(x^*) - \nabla F(x^k)(x^k - x^*)\right\| \leq \varepsilon_k \left\|x^k - x^*\right\| \quad \forall k \geq 0,$$

in other words,

$$\left\|F(x^k) - F(x^*) - \nabla F(x^k)(x^k - x^*)\right\| = o(\left\|x^k - x^*\right\|) \text{ as } k \to +\infty.$$

(b) If F is continuously differentiable and ∇F is locally Lipschitz continuous at x^*, then there exists $C > 0$ such that

$$\left\|F(x^k) - F(x^*) - \nabla F(x^k)(x^k - x^*)\right\| \leq C \left\|x^k - x^*\right\|^2 \quad \forall k \geq 0,$$

in other words,

$$\left\|F(x^k) - F(x^*) - \nabla F(x^k)(x^k - x^*)\right\| = O(\left\|x^k - x^*\right\|^2) \text{ as } k \to +\infty.$$

Proof. Recall that (Fréchet) differentiabilty of F at x^* means

$$\lim_{x\to x^*} \frac{\|F(x) - F(x^*) - \nabla F(x^*)(x - x^*)\|}{\|x - x^*\|} = 0.$$

(a) By the triangle inequality, we have for every $k \geq 0$

$$\begin{aligned} &\left\|F(x^k) - F(x^*) - \nabla F(x^k)(x^k - x^*)\right\| \\ \leq\ &\left\|F(x^k) - F(x^*) - \nabla F(x^*)(x^k - x^*)\right\| + \left\|(\nabla F(x^k) - \nabla F(x^*))(x^k - x^*)\right\|. \end{aligned}$$

We will address the two summands separately. First, we define for every $k \geq 0$

$$\varepsilon'_k := \begin{cases} \frac{\left\|F(x^k) - F(x^*) - \nabla F(x^*)(x^k - x^*)\right\|}{\left\|x^k - x^*\right\|}, & \text{if } x^k \neq x^* \\ 0, & \text{otherwise} \end{cases}.$$

By the differentiability of F at x^* we have that $\varepsilon'_k \downarrow 0$ as $k \to +\infty$. Next, we define for every $k \geq 0$

$$\varepsilon''_k := \left\|\nabla F(x^k) - \nabla F(x^*)\right\|.$$

By the continuity of ∇F at x^*, we have that also $\varepsilon''_k \downarrow 0$ as $k \to +\infty$. We define $\varepsilon_k := \varepsilon'_k + \varepsilon''_k$ for every $k \geq 0$ and get

$$\left\|F(x^k) - F(x^*) - \nabla F(x^k)(x^k - x^*)\right\| \leq \varepsilon_k \left\|x^k - x^*\right\|.$$

(b) Let $B(x^*;\delta)$ be a neighbourhood of x^* on which ∇F is Lipschitz continuous with Lipschitz constant $L_{x^*} \geq 0$. Then there exists $k_0 \geq 0$ be such that $x^k \in B(x^*;\delta)$ for every $k \geq k_0$. Then, by the Mean Value Theorem in integral form, the following holds for every $k \geq k_0$

$$\begin{aligned} & F(x^k) - F(x^*) - \nabla F(x^k)(x^k - x^*) \\ = & \int_0^1 \nabla F(x^* + t(x^k - x^*))(x^k - x^*)\, dt - \nabla F(x^k)(x^k - x^*) \\ = & \int_0^1 \left[\nabla F(x^* + t(x^k - x^*)) - \nabla F(x^k)\right](x^k - x^*)\, dt. \end{aligned}$$

Taking the norm on both sides and invoking the Lipschitz continuity of ∇F on $B(x^*;\delta)$ (note that $x^* + t(x^k - x^*) \in B(x^*;\delta)$ for every $k \geq k_0$), we get for every $k \geq k_0$

$$\begin{aligned} & \|F(x^k) - F(x^*) - \nabla F(x^k)(x^k - x^*)\| \\ \leq & \int_0^1 \left\|\nabla F(x^* + t(x^k - x^*)) - \nabla F(x^k)\right\| \left\|x^k - x^*\right\| \, dt \\ \leq & \int_0^1 L_{x^*}(1-t)\left\|x^k - x^*\right\|^2 \, dt = \frac{L_{x^*}}{2}\left\|x^k - x^*\right\|^2. \end{aligned}$$

This yields that there exists $C > 0$ such that this inequality is fulfilled for every $k \geq 0$. ■

Theorem 8.7 (local convergence theorem of the Newton algorithm) *Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable, x^* a solution of (8.1) and $\nabla F(x^*)$ a regular matrix. Then there exists $\varepsilon > 0$ such that for every $x^0 \in B(x^*;\varepsilon)$ the following statements are true:*

(a) Algorithm 8.3 is well-defined and generates a sequence $(x^k)_{k\geq 0}$ which converges to x^*.

(b) The rate of convergence of $(x^k)_{k\geq 0}$ is superlinear.

(c) If ∇F is locally Lipschitz continuous at x^*, then the rate of convergence of $(x^k)_{k\geq 0}$ is quadratic.

Proof. (a) By Lemma 8.5, we know that there exist $\varepsilon_1 > 0$ and $c > 0$ such that for every $x \in B(x^*;\varepsilon_1)$, the matrix $\nabla F(x)$ is regular and $\|\nabla F(x)^{-1}\| \leq c$. Moreover, since F is differentiable and ∇F is continuous at x^*, we know that there exist $\varepsilon_2 > 0$ such that for every $x \in B(x^*;\varepsilon_2)$ it holds

$$\begin{aligned} & \|F(x) - F(x^*) - \nabla F(x)(x - x^*)\| \\ \leq & \|F(x) - F(x^*) - \nabla F(x^*)(x - x^*)\| + \|\nabla F(x) - \nabla F(x^*)\|\,\|x - x^*\| \\ \leq & \frac{1}{4c}\|x - x^*\| + \frac{1}{4c}\|x - x^*\| = \frac{1}{2c}\|x - x^*\|. \end{aligned}$$

Let $\varepsilon := \min\{\varepsilon_1, \varepsilon_2\}$. For $x^0 \in B(x^*; \varepsilon)$, we know that $\nabla F(x^0)$ is regular and therefore $x^1 := x^0 - \nabla F(x^0)^{-1} F(x^0)$ is well-defined. Furthermore, by using that $F(x^*) = 0$, we have

$$\begin{aligned}
\left\|x^1 - x^*\right\| &= \left\|x^0 - \nabla F(x^0)^{-1} F(x^0) - x^*\right\| \\
&= \left\|\nabla F(x^0)^{-1}(F(x^*) - F(x^0) + \nabla F(x^0)(x^0 - x^*))\right\| \\
&\leq \left\|\nabla F(x^0)^{-1}\right\| \left\|-F(x^*) + F(x^0) - \nabla F(x^0)(x^0 - x^*)\right\| \\
&\leq c \cdot \frac{1}{2c} \left\|x^0 - x^*\right\| = \frac{1}{2} \left\|x^0 - x^*\right\|,
\end{aligned}$$

which implies that $x^1 \in B(x^*; \varepsilon)$ and therefore that x^2 is also well-defined. Repeating this argument for every $k \geq 0$, we get

$$\left\|x^{k+1} - x^*\right\| \leq \frac{1}{2} \left\|x^k - x^*\right\| \leq \left(\frac{1}{2}\right)^k \left\|x^0 - x^*\right\|,$$

which implies that $(x^k)_{k\geq 0} \subseteq B(x^*; \varepsilon)$ and therefore guarantees that the method is well-defined. Lastly, we have that $x^k \to x^*$ as $k \to +\infty$.

(b) We have for every $k \geq 0$

$$\begin{aligned}
\left\|x^{k+1} - x^*\right\| &= \left\|x^k - \nabla F(x^k)^{-1} F(x^k) - x^*\right\| \\
&= \left\|\nabla F(x^k)^{-1}(-F(x^k) + \nabla F(x^k)(x^k - x^*))\right\| \\
&= \left\|\nabla F(x^k)^{-1}(F(x^k) - \nabla F(x^k)(x^k - x^*))\right\| \\
&= \left\|\nabla F(x^k)^{-1}(F(x^k) - F(x^*) - \nabla F(x^k)(x^k - x^*))\right\| \\
&\leq \left\|\nabla F(x^k)^{-1}\right\| \left\|F(x^k) - F(x^*) - \nabla F(x^k)(x^k - x^*)\right\| \\
&\leq c \left\|F(x^k) - F(x^*) - \nabla F(x^k)(x^k - x^*)\right\| \\
&\leq c \cdot \varepsilon_k \cdot \left\|x^k - x^*\right\|,
\end{aligned}$$

where the sequence $(\varepsilon_k)_{k\geq 0} \downarrow 0$ is provided by Lemma 8.6 (a). This shows (b).

(c) The statement follows by making use of Lemma 8.6 (b) instead of Lemma 8.6 (a) in the above estimates. ■

The following examples emphasize the importance of the assumptions in Theorem 8.7.

Example 8.8 (a) The method may fail when F is not continuously differentiable in a neighbourhood of the solution x^*. For instance, take

$$F : \mathbb{R} \to \mathbb{R}, \quad F(x) = x^{1/3}, \quad F'(x) = \frac{1}{3} x^{-2/3} \text{ for } x \neq 0.$$

The function F is not differentiable at 0. Furthermore, its unique zero is $x^* = 0$. For $x^0 \in \mathbb{R} \setminus \{0\}$, we have for every $k \geq 0$

$$x^{k+1} := x^k - \frac{(x^k)^{1/3}}{\frac{1}{3}(x^k)^{-2/3}} = -2x^k = 4x^{k-1} = \ldots = (-2)^{k+1} x^0,$$

therefore $(x^k)_{k\geq 0}$ is not a convergent sequence.

(b) The method may fail when the starting point is not close enough to x^*. For instance, take

$$F : \mathbb{R} \to \mathbb{R}, \quad F(x) = x^3 - 2x + 2, \quad F'(x) = 3x^2 - 2.$$

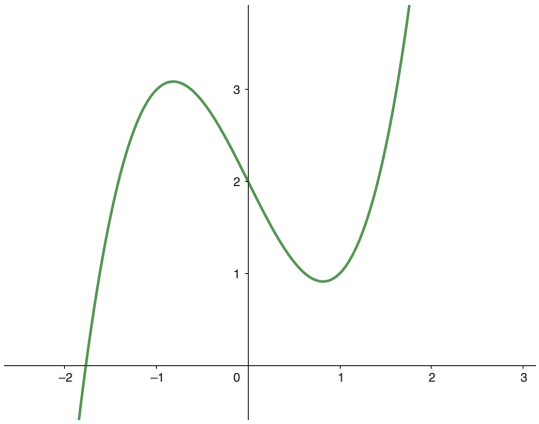

Figure 8.2: The function $F(x) = x^3 - 2x + 2$.

For the unique zero x^* of F, it holds $x^* \in (-2, -1)$ and $F'(x^*) \neq 0$. We have for every $k \geq 0$

$$x^{k+1} := x^k - \frac{(x^k)^3 - 2x^k + 2}{3(x^k)^2 - 2}.$$

Then, for $x^0 = 0$, we have $x^1 = 1$ and $x^2 = 0$, so the method alternates between 0 and 1. In particular, it does not converge to x^*. This happens because x^0 is too far away from x^*.

(c) The convergence is not always quadratic. For instance, take

$$F : \mathbb{R} \to \mathbb{R}, \quad F(x) = x + x^{4/3}, \quad F'(x) = 1 + \frac{4}{3}x^{1/3}.$$

In particular, $F(0) = 0$ and $F'(0) = 1$ and thus, by Theorem 8.7, there exists $\varepsilon > 0$ such that for every $x^0 \in (-\varepsilon, \varepsilon)$ the sequence generated by

$$x^{k+1} = x^k - \frac{x^k + (x^k)^{4/3}}{1 + \frac{4}{3}(x^k)^{1/3}} = \frac{\frac{1}{3}(x^k)^{4/3}}{1 + \frac{4}{3}(x^k)^{1/3}}$$

converges to 0 superlinearly as $k \to +\infty$.

However, if we check the definition of quadratic convergence, we see the following:

$$\frac{|x^{k+1} - 0|}{|x^k - 0|^2} = \frac{1}{3}\left|\frac{(x^k)^{4/3}}{(x^k)^2 + \frac{4}{3}(x^k)^{7/3}}\right| = \frac{1}{3}\left|\frac{1}{(x^k)^{2/3} + \frac{4}{3}x^k}\right| \to +\infty \text{ as } k \to +\infty,$$

since $x^k \to 0$ as $k \to +\infty$. Therefore, the convergence is not quadratic. This happens because F' is not locally Lipschitz continuous at 0. Indeed, there is no $L_0 \geq 0$ and no $\varepsilon > 0$ such that for every $x, y \in (-\varepsilon, \varepsilon)$ it holds

$$|F'(x) - F'(y)| = \frac{4}{3}|x^{1/3} - y^{1/3}| \leq L_0|x - y|.$$

8.3 The Newton algorithm for optimization problems

We consider again the optimization problem (4.1)

$$\min_{x\in\mathbb{R}^n} f(x),$$

but assume this time that $f:\mathbb{R}^n\to\mathbb{R}$ is a twice continuously differentiable function.

The Newton algorithm for solving this problem is nothing else than the Newton algorithm for solving the nonlinear equation

$$\nabla f(x)=0, \tag{8.5}$$

where $\nabla f:\mathbb{R}^n\to\mathbb{R}^n$ denotes the gradient operator of f. The update rule reads for every $k\geq 0$

$$x^{k+1}:=x^k-(\nabla^2 f(x^k))^{-1}\nabla f(x^k),$$

where $\nabla^2 f:\mathbb{R}^n\to\mathbb{R}^{n\times n}$ denotes the Hessian operator of f. In order to avoid the calculation of the inverse of the matrix $\nabla^2 f(x^k)$, one can determine a vector $d^k\in\mathbb{R}^n$ fulfilling the Newton equation

$$\nabla^2 f(x^k)d=-\nabla f(x^k) \tag{8.6}$$

and make the update

$$x^{k+1}:=x^k+d^k. \tag{8.7}$$

Algorithm 8.9 (Newton algorithm for optimization problems)

1: Choose a starting point $x^0\in\mathbb{R}^n$, $\varepsilon\geq 0$ and set $k:=0$.

2: If $\left\|\nabla f(x^k)\right\|\leq\varepsilon$: STOP.

3: Find a solution $d^k\in\mathbb{R}^n$ of the Newton equation

$$\nabla^2 f(x^k)d=-\nabla f(x^k).$$

4: Set $x^{k+1}:=x^k+d^k$, $k:=k+1$ and go to Step 2.

The local convergence theorem follows as a special case of Theorem 8.7. We set $\varepsilon=0$ and assume that Algorithm 8.9 does not terminate after finitely many steps.

Theorem 8.10 (local convergence theorem of the Newton algorithm for optimization problems) *Let $f:\mathbb{R}^n\to\mathbb{R}$ be twice continuously differentiable, $x^*\in\mathbb{R}^n$ a critical point of (4.1), which means that $\nabla f(x^*)=0$, and $\nabla^2 f(x^*)$ a regular matrix. Then there exists $\varepsilon>0$ such that for all $x^0\in B(x^*;\varepsilon)$ the following statements are true:*

(a) Algorithm 8.9 is well-defined and generates a sequence $(x^k)_{k\geq 0}$ which converges to x^*.

(b) The rate of convergence of $(x^k)_{k\geq 0}$ is superlinear.

(c) If $\nabla^2 f$ is locally Lipschitz continuous at x^*, then the rate of convergence of $(x^k)_{k\geq 0}$ is quadratic.

Remark 8.11 (a) To guarantee convergence to the critical point x^*, one has to start in an unknown neighbourhood $B(x^*;\varepsilon)$ of x^*.

(b) The sequence $(x^k)_{k\geq 0}$ might convergence to a local maximum of f, since local maxima also satisfy $\nabla f(x^*) = 0$.

In order to solve these issues, one can combine the fast convergence properties of the Newton method with the "global convergence features" (regarding the choice of x^0) of the gradient method. This is done by first taking gradient steps to enter an appropriate neighbourhood of x^*, and then by taking Newton steps to converge fast.

Algorithm 8.12 (globalized Newton algorithm for optimization problems)

1: Choose a starting point $x^0 \in \mathbb{R}^n$, $\rho > 0$, $p > 2$, $\beta \in (0,1)$, $\sigma \in (0, \frac{1}{2})$, $\varepsilon \geq 0$ and set $k := 0$.

2: If $\left\|\nabla f(x^k)\right\| \leq \varepsilon$: STOP.

3: Find a solution $d^k \in \mathbb{R}^n$ of the Newton equation

$$\nabla^2 f(x^k) d = -\nabla f(x^k). \tag{8.8}$$

If the Newton equation has no solution or if

$$\nabla f(x^k)^T d^k \leq -\rho \left\|d^k\right\|^p \tag{8.9}$$

is not satisfied, then set $d^k := -\nabla f(x^k)$.

4: Find $t_k := \max\{\beta^l : l = 0, 1, \ldots\}$ such that

$$f(x^k + t_k d^k) \leq f(x^k) + \sigma t_k \nabla f(x^k)^T d^k. \tag{8.10}$$

5: Set $x^{k+1} := x^k + t_k d^k$, $k := k + 1$ and go to Step 2.

Remark 8.13 If the Newton equation has no solution or if the solution is "bad", i.e. the inequality

$$\nabla f(x^k)^T d^k \leq -\rho \left\|d^k\right\|^p$$

is not fulfilled, a gradient step is taken instead of a Newton step. Note that (8.9) is not satisfied means that d^k is not a ("good enough") descent direction.

We set $\varepsilon = 0$ and assume that Algorithm 8.12 does not terminate after finitely many steps. First, we prove the following statement.

Theorem 8.14 (subsequence convergence to a critical point) *Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable. Then every accumulation (limit) point of the sequence $(x^k)_{k\geq 0}$ generated by Algorithm 8.12 is a critical point of f.*

Proof. Let $x^* \in \mathbb{R}^n$ be a limit point of $(x^k)_{k\geq 0}$, i.e. there exists a subsequence $(x^{k_l})_{l\geq 0}$ such that $x^{k_l} \to x^*$ as $l \to +\infty$. We assume that $\nabla f(x^*) \neq 0$. For every $k \geq 0$ it holds $\nabla f(x^k)^T d^k < 0$, thus

$$f(x^{k+1}) = f(x^k + t_k d^k) \leq f(x^k) + \sigma t_k \nabla f(x^k)^T d^k < f(x^k),$$

therefore, $(f(x^k))_{k\geq 0}$ is nonincreasing.

If $d^{k_l} = -\nabla f(x_l^k)$ for infinitely many $l \geq 0$, then the conclusion follows from the convergence theorem of the gradient method with Armijo step size rule, Theorem 6.3.

Therefore, we can assume without loss of generality that d^{k_l} is given for all $l \geq 0$ as a solution of the Newton equation (8.8). For all $l \geq 0$ it holds

$$\|\nabla f(x^{k_l})\| = \|\nabla^2 f(x^{k_l}) d^{k_l}\| \leq \|\nabla^2 f(x^{k_l})\| \|d^{k_l}\|. \tag{8.11}$$

We claim that there exist $c_1 > 0$ and $c_2 > 0$ such that

$$0 < c_1 \leq \|d^{k_l}\| \leq c_2 \quad \forall l \geq 0. \tag{8.12}$$

Indeed, assuming that $\inf_{l\geq 0} \|d^{k_l}\| = 0$, there exists a subsequence $(d^{k_{l_s}})_{s\geq 0}$ such that $\|d^{k_{l_s}}\|$ converges to zero as $s \to +\infty$. According to (8.11), this leads to $\nabla f(x^*) = 0$, which is a contradiction. On the other hand, assuming that $\sup_{l\geq 0} \|d^{k_l}\| = +\infty$, there exists a subsequence $(d^{k_{l_s}})_{s\geq 0}$ such that $\|d^{k_{l_s}}\|$ converges to $+\infty$ as $s \to +\infty$. According to (8.9), for every $s \geq 0$ we have

$$\rho \leq \frac{-\nabla f(x^{k_{l_s}})^T d^{k_{l_s}}}{\|d^{k_{l_s}}\|^p} \leq \|\nabla f(x^{k_{l_s}})\| \|d^{k_{l_s}}\|^{1-p},$$

which also leads to a contradiction, since the right-hand side converges to zero as $s \to +\infty$.

Since f is continuous, we have $f(x^{k_l}) \to f(x^*)$ as $l \to +\infty$ and, therefore, $f(x^k) \to f(x^*)$ as $k \to +\infty$. Since for every $k \geq 0$

$$0 \leq \sigma t_k \nabla f(x^k)^T d^k \leq f(x^k) - f(x^{k+1})$$

and $f(x^k) - f(x^{k+1}) \to 0$ as $k \to +\infty$, it yields $t_k \nabla f(x^k)^T d^k \to 0$ as $k \to +\infty$, and so $t_{k_l} \nabla f(x^{k_l})^T d^{k_l} \to 0$ as $l \to +\infty$.

Next we will show that the sequence $(t_{k_l})_{l\geq 0}$ is bounded away from zero. We assome without loss of generality that $t_{k_l} \to 0$ and, in accordance with (8.12), that $d^{k_l} \to d^* \neq 0$ as $l \to +\infty$; otherwise, we may pass to suitable subsequences. For every $l \geq 0$, we have $t_{k_l} := \beta^{m_{k_l}}$. By the Armijo rule, it holds for every $l \geq 0$

$$f(x^{k_l} + \beta^{m_{k_l}-1} d^{k_l}) > f(x^{k_l}) + \sigma \beta^{m_{k_l}-1} \nabla f(x^{k_l})^T d^{k_l}$$

or, equivalently,

$$\frac{f(x^{k_l} + \beta^{m_{k_l}-1} d^{k_l}) - f(x^{k_l})}{\beta^{m_{k_l}-1}} > \sigma \nabla f(x^{k_l})^T d^{k_l}.$$

Note that this holds because, by the Armijo rule, m_{k_l} is the first exponent for which the inequality $f(x^{k_l}+\beta^{m_{k_l}}d^{k_l}) \leq f(x^{k_l})+\sigma\beta^{m_{k_l}}\nabla f(x^{k_l})^T d^{k_l}$ is fulfilled, so for $\beta^{m_{k_l}-1}$, the inequality is not fulfilled. Lemma 6.2 gives

$$\nabla f(x^*)^T d^* \geq \sigma \nabla f(x^*)^T d^*,$$

which implies $\nabla f(x^*)^T d^* \geq 0$. On the other hand, (8.9) implies $\nabla f(x^*)^T d^* \leq -\rho\|d^*\|^p < 0$, which leads to a contradiction.

This means that there exists $\bar{t} > 0$ such that $t_{k_l} \geq \bar{t} > 0$ for all $l \geq 0$, consequently, $\nabla f(x^{k_l})^T d^{k_l} \to 0$ as $l \to +\infty$. Making again use of (8.9), from here it follows that $d^{k_l} \to 0$ as $l \to +\infty$, which is a contradiction to (8.12). In conclusion, $\nabla f(x^*) = 0$. ■

Next, we demonstrate that if one of the accumulation points of the sequence $(x^k)_{k\geq 0}$ possesses a specific property, then the entire sequence converges to this point.

Definition 8.15 (isolated accumulation point) Let $(x^k)_{k\geq 0} \subseteq \mathbb{R}^n$ be a sequence and x^* an accumulation point of it. We say that x^* is an isolated accumulation point of $(x^k)_{k\geq 0}$ if there exists $\varepsilon > 0$ such that $B(x^*;\varepsilon)$ contains no further accumulation points of $(x^k)_{k\geq 0}$.

Lemma 8.16 *Let $x^* \in \mathbb{R}^n$ be an isolated accumulation point of the sequence $(x^k)_{k\geq 0} \subseteq \mathbb{R}^n$ with the property that for any subsequence $(x^{k_l})_{l\geq 0}$ that converges to x^* it holds $\lim_{l\to+\infty}(x^{k_l+1} - x^{k_l}) = 0$. Then the whole sequence $(x^k)_{k\geq 0}$ converges to x^*.*

Proof. We assume that $(x^k)_{k\geq 0}$ does not converge to x^*. Let $\varepsilon > 0$ be such that x^* is the only accumulation point of $(x^k)_{k\geq 0}$ in $\overline{B(x^*;\varepsilon)}$. Let $(x^{k_l})_{l\geq 0}$ be a subsequence of $(x^k)_{k\geq 0}$ such that $x^{k_l} \to x^*$ as $l \to +\infty$ and $(x^{k_l})_{l\geq 0} \subseteq \overline{B(x^*;\varepsilon)}$.

For all $l \geq 0$, let

$$m_l := \max\{t : \|x^s - x^*\| \leq \varepsilon \ \forall k_l \leq s \leq t\}.$$

For all $l \geq 0$, m_l is well-defined, since there exists $t > k_l$ such that $\|x^t - x^*\| > \varepsilon$. This means that for all $l \geq 0$

$$\|x^{m_l} - x^*\| \leq \varepsilon \quad \text{and} \quad \|x^{m_l+1} - x^*\| > \varepsilon.$$

The sequence $(x^{m_l})_{l\geq 0}$ is bounded and each of its accumulation points lies in $\overline{B(x^*;\varepsilon)}$, meaning they are all equal to x^*. This implies that $x^{m_l} \to x^*$ as $l \to +\infty$. Let $l_0 \geq 0$ be such that

$$\|x^{m_l} - x^*\| < \frac{\varepsilon}{2} \quad \forall l \geq l_0.$$

From here, we get that for all $l \geq l_0$

$$\|x^{m_l+1} - x^{m_l}\| \geq \|x^{m_l+1} - x^*\| - \|x^{m_l} - x^*\| > \frac{\varepsilon}{2}.$$

This leads to the desired contradiction. ■

Theorem 8.17 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable, and x^* an isolated accumulation point of the sequence $(x^k)_{k\geq 0}$ generated by Algorithm 8.12. Then $(x^k)_{k\geq 0}$ converges to x^* as $k \to +\infty$.*

Proof. Let $(x^{k_l})_{l\geq 0}$ be a subsequence of $(x^k)_{k\geq 0}$ such that $x^{k_l} \to x^*$ as $l \to +\infty$. According to Theorem 8.14, $\nabla f(x^{k_l}) \to \nabla f(x^*) = 0$ as $l \to +\infty$. Since $t_k \in (0, 1]$, for all $k \geq 0$ it holds

$$\|x^{k+1} - x^k\| = t_k \|d^k\| \leq \|d^k\|.$$

For all $l \geq 0$ for which d^{k_l} fulfills (8.9), by the Cauchy-Schwarz inequality, we have

$$\rho \|d^{k_l}\|^{p-1} \leq \|\nabla f(x^{k_l})\|,$$

while for the other indices $l \geq 0$, we have $d^{k_l} = -\nabla f(x^{k_l})$. This implies that $\|d^{k_l}\| \to 0$ as $l \to +\infty$, therefore, $\|x^{k_l+1} - x^{k_l}\| \to 0$ as $l \to +\infty$. The conclusion follows from the lemma above. ■

The following lemma shows that the positive definiteness of the Hessian of a function f at a point x^* extends uniformly to points in a neighborhood of x^*.

Lemma 8.18 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable, and $x^* \in \mathbb{R}^n$ such that $\nabla^2 f(x^*)$ is positive definite. Then there exist $\delta > 0$ and $\alpha > 0$ such that*

$$d^T \nabla^2 f(x) d \geq \alpha \|d\|^2 \quad \forall x \in B(x^*; \delta) \quad \forall d \in \mathbb{R}^n.$$

Proof. Assume by contradiction that for all $k \geq 1$ there exists $x^k \in B(x^*; \frac{1}{k})$ and $d^k \in \mathbb{R}^n$ such that

$$(d^k)^T \nabla^2 f(x^k) d^k < \frac{1}{k} \|d^k\|^2.$$

This implies that for all $k \geq 1$

$$\left(\frac{d^k}{\|d^k\|}\right)^T \nabla^2 f(x^k) \left(\frac{d^k}{\|d^k\|}\right) < \frac{1}{k}.$$

Considering a subsequence $\left(\frac{d^{k_l}}{\|d^{k_l}\|}\right)_{l\geq 1}$ that converges to $d^* \neq 0$ as $l \to +\infty$, we obtain that $(d^*)^T \nabla^2 f(x^*) d^* \leq 0$, which contradicts the positive definitness of the Hessian. ■

The next result shows that, under certain assumptions, the globalized Newton algorithm accepts the step size $t_k = 1$ for sufficiently large k, provided that the direction d^k is given by a solution of the Newton equation (8.8).

Theorem 8.19 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable, $x^* \in \mathbb{R}^n$ such that $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ is positive definite, $(x^k)_{k\geq 0} \subseteq \mathbb{R}^n$ a sequence which converges to x^* as $k \to +\infty$, and $(d^k)_{k\geq 0} \subseteq \mathbb{R}^n$ the sequence of Newton directions*

$$d^k := -\nabla^2 f(x^k)^{-1} \nabla f(x^k) \quad \forall k \geq 0.$$

If $\sigma \in (0, \frac{1}{2})$, then there exists $k_0 \geq 0$ such that

$$f(x^k + d^k) \leq f(x^k) + \sigma \nabla f(x^k)^T d^k \quad \forall k \geq k_0.$$

Proof. Let $c > 0$ be the constant provided by Lemma 8.5 and $k_0' \geq 0$ such that

$$\|\nabla^2 f(x^k)^{-1}\| \leq c \quad \forall k \geq k_0'.$$

Thus, for all $k \geq k_0'$ it holds

$$\|d^k\| \leq \|\nabla^2 f(x^k)^{-1}\|\|\nabla f(x^k)\| \leq c\|\nabla f(x^k)\|,$$

which proves that $d^k \to 0$ as $k \to +\infty$.

Since $\nabla^2 f(x^*)^{-1}$ is positive definite, according to Lemma 8.18, there exist $\alpha > 0$ and $k_0'' \geq 0$ such that

$$\nabla f(x^k)^T \nabla^2 f(x^k)^{-1} \nabla f(x^k) \geq \alpha \|\nabla f(x^k)\|^2 \quad \forall k \geq k_0''.$$

For all $k \geq 0$, we apply Taylor's Theorem and get $\xi^k \in (x^k, x^k + d^k)$ such that

$$f(x^k + d^k) = f(x^k) + \nabla f(x^k)^T d^k + \frac{1}{2}(d^k)^T \nabla^2 f(\xi^k) d^k.$$

Since $\xi^k \to x^*$ as $k \to +\infty$ and $\sigma \in (0, \frac{1}{2})$, there exists $k_0''' \geq 0$ such that

$$\left(\sigma - \frac{1}{2}\right)\alpha + \frac{1}{2}c^2\|\nabla^2 f(\xi^k) - \nabla^2 f(x^k)\| \leq 0 \quad \forall k \geq k_0'''.$$

Consequently, for all $k \geq k_0 := \max\{k_0', k_0'', k_0'''\}$ it holds

$$\begin{aligned}
f(x^k + d^k) &= f(x^k) + \nabla f(x^k)^T d^k + \frac{1}{2}(d^k)^T \nabla^2 f(\xi^k) d^k \\
&= f(x^k) + \nabla f(x^k)^T d^k + \frac{1}{2}(d^k)^T \nabla^2 f(x^k) d^k + \frac{1}{2}(d^k)^T \left(\nabla^2 f(\xi^k) - \nabla^2 f(x^k)\right) d^k \\
&= f(x^k) + \nabla f(x^k)^T d^k - \frac{1}{2}\nabla f(x^k)^T d^k + \frac{1}{2}(d^k)^T \left(\nabla^2 f(\xi^k) - \nabla^2 f(x^k)\right) d^k \\
&\leq f(x^k) + \frac{1}{2}\nabla f(x^k)^T d^k + \frac{1}{2}\|d^k\|^2 \|\nabla^2 f(\xi^k) - \nabla^2 f(x^k)\| \\
&\leq f(x^k) + \sigma \nabla f(x^k)^T d^k + \left(\frac{1}{2} - \sigma\right)\nabla f(x^k)^T d^k + \frac{1}{2}\|d^k\|^2 \|\nabla^2 f(\xi^k) - \nabla^2 f(x^k)\| \\
&= f(x^k) + \sigma \nabla f(x^k)^T d^k - \left(\frac{1}{2} - \sigma\right)\nabla f(x^k)^T \nabla^2 f(x^k)^{-1} \nabla f(x^k) \\
&\quad + \frac{1}{2}\|d^k\|^2 \|\nabla^2 f(\xi^k) - \nabla^2 f(x^k)\| \\
&\leq f(x^k) + \sigma \nabla f(x^k)^T d^k - \left(\frac{1}{2} - \sigma\right)\alpha \|\nabla f(x^k)\|^2 + \frac{1}{2}\|d^k\|^2 \|\nabla^2 f(\xi^k) - \nabla^2 f(x^k)\| \\
&\leq f(x^k) + \sigma \nabla f(x^k)^T d^k + \left(\left(\sigma - \frac{1}{2}\right)\alpha + \frac{1}{2}c^2\|\nabla^2 f(\xi^k) - \nabla^2 f(x^k)\|\right)\|\nabla f(x^k)\|^2 \\
&\leq f(x^k) + \sigma \nabla f(x^k)^T d^k,
\end{aligned}$$

which concludes the proof. ■

We are now ready to present the following convergence theorem for sequences $(x^k)_{k\geq 0}$ generated by the globalized Newton algorithm that have an accumulation point $x^* \in \mathbb{R}^n$ for which $\nabla^2 f(x^*)$ is positive definite.

Theorem 8.20 *Let $(x^k)_{k\geq 0}$ be a sequence generated by Algorithm 8.12 and x^* be an accumulation (limit) point of $(x^k)_{k\geq 0}$ such that $\nabla^2 f(x^*)$ is positive definite. Then the following statements are true:*

(a) The entire sequence $(x^k)_{k\geq 0}$ converges to x^* as $k \to +\infty$, and x^* is a strict local minimum of f.

(b) For sufficiently large k, the direction d^k is the solution of the Newton equation (8.8).

(c) For sufficiently large k, the algorithm accepts the step size $t_k = 1$.

(d) The sequence $(x^k)_{k\geq 0}$ converges superlinearly to x^* as $k \to +\infty$.

(e) If, in addition, $\nabla^2 f$ is locally Lipschitz continuous at x^*, then $(x^k)_{k\geq 0}$ converges quadratically to x^* as $k \to +\infty$.

Proof. (a) According to Theorem 8.14, x^* is a critical point of f. Since $(f(x^k))_{k\geq 0}$ is nonincreasing and a subsequence of it converges to $f(x^*)$, $f(x^k) \to f(x^*)$ as $k \to +\infty$. This shows that f takes at every accumulation point of $(x^k)_{k\geq 0}$ the value $f(x^*)$. Since $\nabla^2 f(x^*)$ is positive definite, according to Theorem 3.3, x^* is a strict local minimum of f. This proves that x^* is an isolated accumulation point of $(x^k)_{k\geq 0}$. By Theorem 8.17, $x^k \to x^*$ as $k \to +\infty$.

(b) Since $\nabla^2 f(x^*)$ is positive definite, there exists $k_0 \geq 0$ such that the matrices $\nabla^2 f(x^k)$ are positive definite and therefore invertible. This means that the Newton equation has for all $k \geq k_0$ a unique solution. By Lemma 8.5, there exist $c > 0$ and $k_1 \geq k_0$ such that for all $k \geq k_1$

$$\|\nabla^2 f(x^k)^{-1}\| \leq c,$$

consequently,

$$\|d^k\| \leq c\|\nabla f(x^k)\|.$$

Applying again Lemma 8.18, there exist $\alpha > 0$ and $k_2 \geq k_1$ such that for all $k \geq k_2$

$$d^T \nabla^2 f(x^k)^{-1} d \geq \alpha \|d\|^2 \quad \forall d \in \mathbb{R}^n.$$

Therefore, for all $k \geq k_2$ it holds

$$\frac{\alpha}{c^2}\|d^k\|^2 = \frac{\alpha}{c^2}\|\nabla^2 f(x^k)^{-1}\nabla f(x^k)\|^2 \leq \alpha\|\nabla f(x^k)\|^2 \leq \nabla f(x^k)^T \nabla^2 f(x^k)^{-1}\nabla f(x^k) = -\nabla f(x^k)^T d^k.$$

From $\nabla f(x^k) \to \nabla f(x^*) = 0$ as $k \to +\infty$, it yields $\|d^k\| \to 0$ as $k \to +\infty$. This yields that there exists $k_3 \geq k_2$ such that for all $k \geq k_3$

$$\nabla f(x^k)^T d^k \leq -\frac{\alpha}{c^2}\|d^k\|^2 \leq -\rho\|d^k\|^p,$$

which proves (b).

(c) The statement is a direct consequence of (a), (b), and Theorem 8.19.

(d)-(e) Due to (a), (b) and (c), Algorithm 8.12 coincides in a neighbourhood of x^* with the local Newton algorithm. Consequently, the two algorithms share the same convergence properties. The statements follow from Theorem 8.10 (b) and (c), respectively. ■

Chapter III

Numerical methods for constrained optimization problems

In this chapter we will study numerical methods for optimization problems of the form (1.5)

$$\begin{aligned} &\min f(x). \\ \text{such that } &g_i(x) \leq 0, i = 1, \ldots, m \\ &h_j(x) = 0, i = 1, \ldots, p \\ &x \in \mathbb{R}^n \end{aligned}$$

9 Penalty methods

9.1 The penalty algorithm

The constrained optimization problem (1.5) will be approached by a sequence of unconstrained optimization problems. In each iteration, an unconstrained optimization problem is solved. The objective function of every unconstrained problem will penalize the violation of the constraints of (1.5). We will first consider optimization problems with only equality constraints, more specifically,

$$\begin{aligned} &\min f(x), \\ \text{such that } &h_j(x) = 0, i = 1, \ldots, p \\ &x \in \mathbb{R}^n \end{aligned} \tag{9.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $h = (h_1, \ldots, h_p) : \mathbb{R}^n \to \mathbb{R}^p$ are continuous functions.

Definition 9.1 (penalty function) Let $\alpha > 0$. The function

$$P(\cdot; \alpha) : \mathbb{R}^n \to \mathbb{R}, \quad P(x; \alpha) = f(x) + \frac{\alpha}{2} \|h(x)\|^2,$$

is called penalty function. In this context, $\alpha > 0$ is called penalty parameter.

Remark 9.2 Let $X := \{x \in \mathbb{R}^n \mid h_j(x) = 0, j = 1, \ldots, p\}$ denote the feasible set of the problem (9.1). Then, for all $\alpha > 0$, $P(\cdot; \alpha)$ and f coincide on X.

The idea of a penalty algorithm is to minimize the penalty function $P(\cdot; \alpha)$ in order to get a solution of (9.1). In particular, we want to force the element to be feasible. To this end we will take α large, which forces $\|h(\cdot)\|^2$ to be small.

Example 9.3 Let $f, h : \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$ and $h(x) = x - 1$. The global minimum of (9.1) in this case is $x^* = 1$. We calculate $x^*(\alpha)$ as the global minimum of $P(\cdot; \alpha)$ over $\mathbb{R}$ for growing values of α:

$$\begin{aligned}
&\alpha = 1 \;\Rightarrow\; P(x;1) = \frac{3}{2}x^2 - x + 1 \;\Rightarrow\; x^*(1) = \frac{1}{3}\\
&\alpha = 10 \;\Rightarrow\; P(x;10) = 6x^2 - 10x + 1 \;\Rightarrow\; x^*(10) = \frac{5}{6}\\
&\alpha = 20 \;\Rightarrow\; P(x;20) = 11x^2 - 20x + 1 \;\Rightarrow\; x^*(20) = \frac{10}{11}\\
&\vdots \qquad\qquad \vdots \qquad\qquad \vdots \qquad\qquad \vdots\\
&\alpha = k \;\Rightarrow\; P(x;k) = \frac{k+2}{2}x^2 - kx + 1 \;\Rightarrow\; x^*(k) = \frac{k}{k+2}.
\end{aligned}$$

We see that $x^*(\alpha) \to 1$ as $\alpha \to +\infty$.

For a strictly monotonically increasing sequence $(\alpha_k)_{k\geq 0} \subseteq (0, +\infty)$, we calculate x^k as a "minimum" of $P(x; \alpha_k)$. We hope that $(x^k)_{k\geq 0}$ converges to a global minimum of (9.1).

Algorithm 9.4 (The penalty algorithm)

1: Choose $\alpha_0 > 0$ and set $k := 0$.

2: Find x^k as a global minimum of the unconstrained optimization problem

$$\min_{x\in\mathbb{R}^n} P(x; \alpha_k). \tag{9.2}$$

3: If $h(x^k) = 0$: STOP.

4: Choose $\alpha_{k+1} > \alpha_k$, set $k := k + 1$ and go to Step 2.

Theorem 9.5 *Let $f : \mathbb{R}^n \to \mathbb{R}$ and $h = (h_1, \ldots, h_p) : \mathbb{R}^n \to \mathbb{R}^p$ be continuous functions, and $(\alpha_k)_{k\geq 0}$ a strictly monotonically increasing sequence with $\alpha_k \to +\infty$ as $k \to +\infty$. Assume that the feasible set X is nonempty and denote $f^* := \inf_{x\in X} f(x) < +\infty$. If $(x^k)_{k\geq 0}$ is the sequence generated by Algorithm 9.4, then the following statements are true:*

(a) The sequence $(P(x^k; \alpha_k))_{k\geq 0}$ is monotonically increasing.

(b) The sequence $(\|h(x^k)\|)_{k\geq 0}$ is monotonically decreasing.

(c) The sequence $(f(x^k))_{k\geq 0}$ is monotonically increasing.

(d) It holds $h(x^k) \to 0$ as $k \to +\infty$.

(e) Every accumulation point of $(x^k)_{k\geq 0}$ is a global minimum of (9.1).

Proof. Let $k \geq 0$.

(a) Since x^k minimizes $P(\cdot;\alpha_k)$, we have $P(x^k;\alpha_k) \leq P(x^{k+1};\alpha_k)$. In addition, it holds

$$\begin{aligned} P(x^{k+1};\alpha_k) &= f(x^{k+1}) + \frac{\alpha_k}{2}\left\|h(x^{k+1})\right\|^2 \\ &\leq f(x^{k+1}) + \frac{\alpha_{k+1}}{2}\left\|h(x^{k+1})\right\|^2 = P(x^{k+1};\alpha_{k+1}). \end{aligned}$$

Therefore,

$$P(x^k;\alpha_k) \leq P(x^{k+1};\alpha_k) \leq P(x^{k+1};\alpha_{k+1}).$$

(b) We have

$$P(x^k;\alpha_k) + P(x^{k+1};\alpha_{k+1}) \leq P(x^{k+1};\alpha_k) + P(x^k;\alpha_{k+1})$$

or, equivalently,

$$\frac{\alpha_k}{2}\left\|h(x^k)\right\|^2 + \frac{\alpha_{k+1}}{2}\left\|h(x^{k+1})\right\|^2 \leq \frac{\alpha_k}{2}\left\|h(x^{k+1})\right\|^2 + \frac{\alpha_{k+1}}{2}\left\|h(x^k)\right\|^2.$$

This is further equivalent to

$$(\alpha_{k+1}-\alpha_k)(\left\|h(x^{k+1})\right\|^2 - \left\|h(x^k)\right\|^2) \leq 0 \;\Leftrightarrow\; \left\|h(x^{k+1})\right\|^2 \leq \left\|h(x^k)\right\|^2,$$

because $\alpha_{k+1} - \alpha_k > 0$.

(c) Using $P(x^k;\alpha_k) \leq P(x^{k+1};\alpha_k)$ and part (b), we get

$$f(x^k) + \frac{\alpha_k}{2}\left\|h(x^k)\right\|^2 \leq f(x^{k+1}) + \frac{\alpha_k}{2}\left\|h(x^{k+1})\right\|^2 \leq f(x^{k+1}) + \frac{\alpha_k}{2}\left\|h(x^k)\right\|^2,$$

therefore $f(x^k) \leq f(x^{k+1})$.

(d) We have

$$P(x^k;\alpha_k) = \inf_{x\in\mathbb{R}^n} P(x;\alpha_k) \leq \inf_{x\in X} P(x;\alpha_k) = \inf_{x\in X} f(x) = f^* < +\infty. \tag{9.3}$$

On the other hand, by (c),

$$P(x^k;\alpha_k) = f(x^k) + \frac{\alpha_k}{2}\left\|h(x^k)\right\|^2 \geq f(x^0) + \frac{\alpha_k}{2}\left\|h(x^k)\right\|^2 > f(x^0).$$

Since $\alpha_k \to +\infty$ as $k \to +\infty$, the statement follows.

(e) Let x^* be a limit point of $(x^k)_{k\geq 0}$. Then there exists a subsequence $x^{k_l} \to x^*$ as $l \to +\infty$, and therefore by continuity of h and (d), it holds $h(x^*) = 0$, in other words $x^* \in X$. Furthermore, by (9.3),

$$f(x^*) = \lim_{l\to+\infty} f(x^{k_l}) \leq \lim_{l\to\infty} P(x^{k_l};\alpha_{k_l}) \leq f^* = \inf_{x\in X} f(x),$$

and thus x^* is a global minimum for (9.1). ■

Remark 9.6 The stopping criterion in Algorithm 9.4 makes sense. Indeed, inequality (9.3) implies that for all $k \geq 0$

$$f(x^k) \leq P(x^k; \alpha_k) \leq f^* = \inf_{x \in X} f(x).$$

If $h(x^k) = 0$ for an index $k \geq 0$, we have $x^k \in X$, which means that x^k is a minimum for (9.1).

Remark 9.7 In order to obtain a similar algorithm with the corresponding convergence statement for the optimization problem (1.5), we just have to reformulate this as

$$\begin{aligned} &\min f(x), \\ \text{such that } &(g_i)_+(x) := \max\{0, g_i(x)\} = 0, i = 1, \ldots, m \\ &h_j(x) = 0, j = 1, \ldots, p, \\ &x \in \mathbb{R}^n. \end{aligned}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $g = (g_1, \ldots, g_m) : \mathbb{R}^n \to \mathbb{R}^m$ and $h = (h_1, \ldots, h_p) : \mathbb{R}^n \to \mathbb{R}^p$ are continuous functions. The corresponding penalty function reads for $\alpha > 0$

$$\begin{aligned} P(x; \alpha) &= f(x) + \frac{\alpha}{2} \|h(x)\|^2 + \frac{\alpha}{2} \|g_+(x)\|^2 \\ &= f(x) + \frac{\alpha}{2} \|h(x)\|^2 + \frac{\alpha}{2} \sum_{i=1}^{m} (\max\{0, g_i(x)\})^2. \end{aligned}$$

By using the "max function" we might lose differentiability. However, in this chapter we only need the continuity of g, so this is not a problem for now. Note that ultimately we want to minimize $P(\cdot; \alpha)$ using an algorithm that definitely depends on the differentiability properties of h and g. In particular, even if h and g are twice continuously differentiable, $P(\cdot; \alpha)$ may only be once continuously differentiable.

Next, we will assume that f and h in (9.1) are continuously differentiable. If x^k is a minimizer of $P(\cdot; \alpha_k)$, for $k \geq 0$, then

$$0 = \nabla P(x^k; \alpha_k) = \nabla f(x^k) + \alpha_k \sum_{j=1}^{p} h_j(x^k) \nabla h_j(x^k). \tag{9.4}$$

Choosing

$$\mu_j^k := \alpha_k h_j(x^k) \in \mathbb{R}, \ j = 1, \ldots, p, \quad \text{and} \quad \mu^k := (\mu_1^k, \ldots, \mu_p^k)^T \in \mathbb{R}^p, \tag{9.5}$$

one may expect that the sequence $(x^k, \mu^k)_{k \geq 0}$ has as a limit point a KKT point (x^*, μ^*) of (9.1).

Theorem 9.8 *Let $f : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}^p$ be continuously differentiable functions. Let $(x^k)_{k \geq 0}$ be a sequence generated by Algorithm 9.4 with $\lim_{k \to +\infty} x^k = x^*$ such that the vectors $\nabla h_1(x^*), \ldots, \nabla h_p(x^*)$ are linearly independent, i.e. (LICQ) for (9.1) is fulfilled at x^*. Let $(\mu^k)_{k \geq 0}$ be the sequence defined in (9.5). Then, the following statements are true:*

(a) The sequence $(\mu^k)_{k\geq 0}$ converges to an element $\mu^* \in \mathbb{R}^p$ as $k \to +\infty$.

(b) The pair (x^*, μ^*) is a KKT point of (9.1), where μ^* is the uniquely determined Lagrange multiplier corresponding to x^*.

Proof. (a) For all $k \geq 0$, let $A_k := \nabla h(x^k) \in \mathbb{R}^{p\times n}$ and $A_* := \nabla h(x^*) \in \mathbb{R}^{p\times n}$. Then, as h is continuously differentiable, we have

$$A_k \to A_* \quad \text{and} \quad A_k A_k^T \to A_* A_*^T \quad \text{as} \quad k \to +\infty.$$

The matrix $A_* A_*^T$ is positive definite, therefore invertible. Indeed, for all $z \in \mathbb{R}^p$ it holds

$$z^T A_* A_*^T z = (A_*^T z)^T A_*^T z = \left\|A_*^T z\right\|^2 \geq 0,$$

and furthermore

$$\left\|A_*^T z\right\|^2 = 0 \;\Leftrightarrow\; A_*^T z = 0 \;\Leftrightarrow\; \sum_{i=1}^p z_i \nabla h_i(x^*) = 0 \;\Leftrightarrow\; z = 0,$$

where the last equivalence follows by (LICQ).

Let $k_0 \geq 0$ be such that for all $k \geq k_0$ the matrix $A_k A_k^T$ is regular and $(A_k A_k^T)^{-1} \to (A_* A_*^T)^{-1}$ as $k \to +\infty$. Combining (9.4) and (9.5), we obtain for all $k \geq k_0$

$$0 = \nabla f(x^k) + \sum_{j=1}^p \mu_j^k \nabla h_j(x^k) \Leftrightarrow \sum_{j=1}^p \mu_j^k \nabla h_j(x^k) = -\nabla f(x^k) \Leftrightarrow A_k^T \mu^k = -\nabla f(x^k).$$

Therefore, for all $k \geq k_0$,

$$A_k A_k^T \mu^k = -A_k \nabla f(x^k) \;\Leftrightarrow\; \mu^k = -(A_k A_k^T)^{-1} A_k \nabla f(x^k).$$

Taking the limit as $k \to +\infty$, it yields

$$\mu^k \to -(A_* A_*^T)^{-1} A_* \nabla f(x^*) =: \mu^*.$$

(b) For all $k \geq k_0$ it holds

$$0 = \nabla f(x^k) + \sum_{j=1}^p \mu_j^k \nabla h_j(x^k).$$

Taking the limit as $k \to +\infty$, we get

$$0 = \nabla f(x^*) + \sum_{j=1}^p \mu_j^* \nabla h_j(x^*).$$

In addition, we have by Theorem 9.5

$$h(x^*) = \lim_{k\to\infty} h(x^k) = 0.$$

Therefore, (x^*, μ^*) is a KKT point of (9.1), and the uniqueness of μ^* follows from Theorem 2.8. ■

9.2 Exact penalization

The idea of exact penalization is to determine a fixed "convenient" penalty parameter $\bar{\alpha}$ instead of a sequence of penalty parameters and to consequently solve only one unconstrained problem. Consider the constrained problem (1.5)

$$\begin{aligned} &\min f(x), \\ \text{such that } &g_i(x) \leq 0, i = 1, \ldots, m \\ &h_j(x) = 0, i = 1, \ldots, p \\ &x \in \mathbb{R}^n \end{aligned}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $g = (g_1, \ldots, g_m) : \mathbb{R}^n \to \mathbb{R}^m$ and $h = (h_1, \ldots, h_p) : \mathbb{R}^n \to \mathbb{R}^p$ are continuously differentiable functions. One can construct a general class of penalty functions by choosing

$$P_r(x; \alpha) = f(x) + \alpha r(x), \tag{9.6}$$

where $r : \mathbb{R}^n \to \mathbb{R}$ is a function, which is at least continuous, such that

$$r(x) \geq 0 \ \ \forall x \in \mathbb{R}^n$$

and

$$r(x) = 0 \ \Leftrightarrow \ x \in X,$$

where X is the feasible set of (1.5). For

$$r(x) := \frac{1}{2} \left\| (g_+(x), h(x)) \right\|^2,$$

we obtain the penalty function from Remark 9.7.

Definition 9.9 A penalty function of the form (9.6) is called exact at a local minimum x^* of (1.5) if there exists $\bar{\alpha} > 0$ such that x^* is a local minimum of $P_r(\cdot; \alpha)$ for all $\alpha \geq \bar{\alpha}$.

We will show that if P_r is exact at a local minimum x^*, then r cannot be differentiable. This means that one cannot a priori use algorithms designed to solve unconstrained differentiable optimization problems to minimize $P_r(\cdot; \alpha)$.

Theorem 9.10 *Let x^* be a local minimum of (1.5) such that $\nabla f(x^*) \neq 0$. Let P_r be exact at x^*. Then r is not differentiable at x^*.*

Proof. If r is differentiable at x^*, then there exists $\bar{\alpha}$ such that for all $\alpha \geq \bar{\alpha}$, it holds

$$\nabla P_r(x^*; \bar{\alpha}) = 0 \ \Leftrightarrow \ \nabla f(x^*) + \alpha \nabla r(x^*) = 0.$$

This implies that for all $\alpha_1, \alpha_2 \geq \bar{\alpha}$ with $\alpha_1 \neq \alpha_2$, we have

$$\alpha_1 \nabla r(x^*) = -\nabla f(x^*) = \alpha_2 \nabla r(x^*),$$

which implies $\nabla r(x^*) = 0$ and thus $\nabla f(x^*) = 0$ – a contradiction. ■

Remark 9.11 The assumption that $\nabla f(x^*) \neq 0$ in Theorem 9.10 is necessary, however, it is not very restrictive. If x^* is a local minimum of (1.5), we usually do not expect that $\nabla f(x^*) = 0$. It happens if $x^* \in \text{int}(X)$, however, in this situation we do not need the penalty method at all – we can use algorithms for unconstrained optimization problems to find a local minimum of f.

All this motivates the ansatz

$$r(x) = \|(g_+(x), h(x))\|_q ,$$

where

$$\|z\|_q = \begin{cases} (\sum_i |z_i|^q)^{1/q}, & \text{if } 1 \leq q < \infty, \\ \max_i\{|z_i|\}, & \text{if } q = \infty, \end{cases}$$

which leads to

$$P_q(x;\alpha) = f(x) + \alpha \|(g_+(x), h(x))\|_q . \tag{9.7}$$

For $q = 1$, we obtain the exact ℓ_1 penalty function

$$P_1(x;\alpha) = f(x) + \alpha \sum_{i=1}^{m} \max\{0, g_i(x)\} + \alpha \sum_{j=1}^{p} |h_j(x)|.$$

For $q = \infty$, we obtain the exact ℓ_∞ penalty function

$$P_\infty(x;\alpha) = f(x) + \alpha \max\left\{\max\{0, g_1(x)\}, ..., \max\{0, g_m(x)\}, |h_1(x)|, ..., |h_p(x)|\right\}.$$

For $q = 2$, we obtain the exact ℓ_2 penalty function

$$P_2(x;\alpha) = f(x) + \alpha \left(\sum_{i=1}^{m} (\max\{0, g_i(x)\})^2 + \sum_{j=1}^{p} (h_j(x))^2 \right)^{1/2}.$$

Theorem 9.12 *Let $q \in [1, \infty]$ be such that P_q is exact at a local minimum x^* of (1.5). Then $P_{q'}$ is exact at x^* for every $q' \in [1, \infty]$.*

Proof. Assume that P_q is exact at x^* for $q \in [1, \infty]$. In other words, there exists $\bar{\alpha} > 0$ such that for all $\alpha > \bar{\alpha}$, x^* is a local minimum for $P_q(\cdot;\alpha)$. This means that for all $\alpha > \bar{\alpha}$ there exists $\varepsilon(\alpha)$ such that for all $x \in B(x^*;\varepsilon(\alpha))$ it holds

$$P_q(x^*;\alpha) \leq P_q(x;\alpha).$$

Let $q' \in [1, \infty]$. Then there exist $c_1, c_2 > 0$ such that for all $z \in \mathbb{R}^{m+p}$

$$c_1 \|z\|_{q'} \leq \|z\|_q \leq c_2 \|z\|_{q'} .$$

In particular, for $z = (1, 0, \ldots, 0)^T$, we get

$$c_1 \leq 1 \leq c_2.$$

Let $\bar{\alpha}' := c_2\bar{\alpha}$ and take $\alpha \geq \bar{\alpha}'$. Note that $\frac{\alpha}{c_2} \geq \frac{\bar{\alpha}'}{c_2} = \bar{\alpha}$. Then, for all $x \in B(x^*; \varepsilon(\alpha/c_2))$, it holds

$$P_{q'}(x^*; \alpha) = f(x^*) + \alpha \left\| (g_+(x^*), h(x^*)) \right\|_{q'} = f(x^*).$$

For this same reason, we get for all $x \in B(x^*; \varepsilon(\alpha/c_2))$

$$\begin{aligned} P_{q'}(x^*; \alpha) = f(x^*) &= f(x^*) + \frac{\alpha}{c_2} \left\| (g_+(x^*), h(x^*)) \right\|_q \\ &= P_q(x^*; \alpha/c_2) \leq P_q(x; \alpha/c_2) \\ &= f(x) + \frac{\alpha}{c_2} \left\| (g_+(x), h(x)) \right\|_q \leq f(x) + \alpha \left\| (g_+(x), h(x)) \right\|_{q'} \\ &= P_{q'}(x; \alpha). \end{aligned}$$

Therefore, x^* is a local minimum of $P_{q'}(\cdot; \alpha)$. ■

Next, we will show that, for some classes of problems, exact penalty functions can be constructed.

Theorem 9.13 *Let $(x^*, \lambda^*, \mu^*) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$ be a KKT point of the convex optimization problem* (2.1)

$$\begin{aligned} &\min f(x), \\ \text{such that } &g_i(x) \leq 0, \ i = 1, \ldots, m \\ &Ax = b \\ &x \in \mathbb{R}^n \end{aligned}$$

where $f, g_i : \mathbb{R}^n \to \mathbb{R}, i = 1, ..., m$, are continuously differentiable and convex functions, and $h : \mathbb{R}^n \to \mathbb{R}^p, h(x) = Ax - b$. Then there exists $\bar{\alpha} > 0$ such that x^* is a global minimum of $P_1(\cdot; \alpha)$ for all $\alpha \geq \bar{\alpha}$. In other words, P_1 is exact at x^*.

Proof. First, notice that

$$x \mapsto L(x, \lambda^*, \mu^*) = f(x) + \sum_{i=1}^{m} \lambda_i^* g_i(x) + \sum_{j=1}^{p} \mu_j^* h_j(x)$$

is convex. The gradient inequality, together with the fact that (x^*, λ^*, μ^*) is a KKT point, implies that for all $x \in \mathbb{R}^n$ it holds

$$0 = \nabla_x L(x^*, \lambda^*, \mu^*)^T (x - x^*) \leq L(x, \lambda^*, \mu^*) - L(x^*, \lambda^*, \mu^*)$$

or, equivalently,

$$L(x^*, \lambda^*, \mu^*) \leq L(x, \lambda^*, \mu^*).$$

We define

$$\bar{\alpha} := \left\| (\lambda^*, \mu^*) \right\|_\infty = \max\{\lambda_1^*, \ldots, \lambda_m^*, |\mu_1^*|, \ldots, |\mu_p^*|\},$$

and choose $\alpha \geq \bar{\alpha}$. Then, for all $x \in \mathbb{R}^n$ we have

$$\begin{aligned} P_1(x^*;\alpha) &= f(x^*) + \alpha \left\| (g_+(x^*), h(x^*)) \right\|_1 \\ &= f(x^*) \\ &= f(x^*) + \sum_{i=1}^{m} \lambda_i^* g_i(x^*) + \sum_{j=1}^{p} \mu_j^* h_j(x^*). \end{aligned}$$

Continuing from here, we get for all $x \in \mathbb{R}^n$

$$\begin{aligned} P_1(x^*;\alpha) &= L(x^*, \lambda^*, \mu^*) \leq L(x, \lambda^*, \mu^*) \\ &= f(x) + \sum_{i=1}^{m} \lambda_i^* g_i(x) + \sum_{j=1}^{p} \mu_j^* h_j(x) \\ &\leq f(x) + \sum_{i=1}^{m} \lambda_i^* \max\{0, g_i(x)\} + \sum_{j=1}^{p} |\mu_j^*||h_j(x)| \\ &\leq f(x) + \bar{\alpha}\left(\sum_{i=1}^{m} \max\{0, g_i(x)\} + \sum_{j=1}^{p} |h_j(x)| \right) \\ &\leq f(x) + \bar{\alpha} \left\| (g_+(x), h(x)) \right\|_1 \\ &= P_1(x;\alpha). \end{aligned}$$

■

Remark 9.14 According to Theorem 9.13, if x^* is a global minimum of (2.1) and (Slater CQ) is fulfilled, then P_q is exact at x^* for all $q \in [1, \infty]$.

We close this section by addressing the exactness of the ℓ_2 penalty function P_2 in the context of the optimization problem (1.5) with $f, g_i, h_j : \mathbb{R}^n \to \mathbb{R}, i = 1, ..., m, j = 1, ..., p$, continuously differentiable functions. Recall that for all $x \in \mathbb{R}^n$

$$P_2(x;\alpha) = f(x) + \alpha\left(\sum_{i=1}^{m} (\max\{0, g_i(x)\})^2 + \sum_{j=1}^{p} (h_j(x))^2 \right)^{1/2}$$

and

$$r_2(x) := \left(\sum_{i=1}^{m} (\max\{0, g_i(x)\})^2 + \sum_{j=1}^{p} (h_j(x))^2 \right)^{1/2}.$$

Lemma 9.15 *Let x^* be a local minimum of (1.5), which fulfills (MFCQ). Then, for every sequence $(x^k)_{k\geq 0}$ with $x^k \to x^*$ as $k \to +\infty$ and $x^k \notin X$ for all $k \geq 0$, there exists $c > 0$ such that $\left\| \nabla r_2(x^k) \right\| \geq c$ for all $k \geq 0$.*

Proof. Let $(x^k)_{k\geq 0}$ be a sequence such that $x^k \to x^*$ as $k \to +\infty$, $x^k \notin X$ for all $k \geq 0$, and $\left\|\nabla r_2(x^k)\right\| \to 0$ as $k \to +\infty$. Note that if $x^k \notin X$, then $r_2(x^k) > 0$ and therefore, $\nabla r_2(x^k)$ exists. An easy calculation yields

$$\begin{aligned}\nabla r_2(x^k) &= \frac{\sum_{i=1}^m \max\{0, g_i(x^k)\}\nabla g_i(x^k) + \sum_{j=1}^p h_j(x^k)\nabla h_j(x^k)}{r_2(x^k)} \\ &= \sum_{i=1}^m \rho_i^k \nabla g_i(x^k) + \sum_{j=1}^p \mu_j^k \nabla h_j(x^k),\end{aligned}$$

where for all $k \geq 0$ we define

$$\begin{aligned}\rho_i^k &:= \frac{1}{r_2(x^k)} \max\{0, g_i(x^k)\},\ i = 1, ..., m, \\ \mu_j^k &:= \frac{1}{r_2(x^k)} h_j(x^k),\ j = 1, ..., p.\end{aligned}$$

For all $k \geq 0$ it holds

$$\left\|(\rho^k, \mu^k)\right\|_2 = 1,$$

where $\rho^k := (\rho_1^k, \ldots, \rho_m^k)^T \in \mathbb{R}^m$ and $\mu^k := (\mu_1^k, \ldots, \mu_p^k)^T \in \mathbb{R}^p$. Then there exists a subsequence $(\rho^{k_l}, \mu^{k_l})_{l\geq 0}$ such that

$$\lim_{l\to+\infty} (\rho^{k_l}, \mu^{k_l}) = (\rho, \mu) \in \mathbb{R}^m \times \mathbb{R}^n \text{ and } \|(\rho, \mu)\|_2 = 1.$$

Obviously, $\rho_i \geq 0$ for all $i = 1, \ldots, m$. We will prove that actually $\rho_i = 0$ for all $i = 1, \ldots, m$.

As a first step, let $i \notin \mathcal{A}(x^*)$. Then

$$0 > g_i(x^*) = \lim_{k\to\infty} g_i(x^k),$$

so for k large enough, we have $g_i(x^k) < 0$ and therefore, by definition of ρ_i^k,

$$\rho_i = \lim_{l\to+\infty} \rho_i^{k_l} = 0.$$

Next, we have for all $l \geq 0$

$$\nabla r_2(x^{k_l}) = \sum_{i=1}^m \rho_i^{k_l} \nabla g_i(x^{k_l}) + \sum_{j=1}^p \mu_j^{k_l} \nabla h_j(x^{k_l}).$$

Taking the limit as $l \to +\infty$, we get

$$0 = \sum_{i=1}^m \rho_i \nabla g_i(x^*) + \sum_{j=1}^p \mu_j \nabla h_j(x^*). \tag{9.8}$$

Multiplying this equation with the vector $d \in \mathbb{R}^n$ given by (MFCQ) (b), we get

$$0 = \sum_{i \in \mathcal{A}(x^*)} \rho_i \nabla g_i(x^*)^T d + \sum_{j=1}^{p} \mu_j \nabla h_j(x^*)^T d.$$

Since $\nabla g_i(x^*)^T d < 0$ for all $i \in \mathcal{A}(x^*)$ and $\nabla h_j(x^*)^T d = 0$ for all $j = 1, ...p$, we conclude that $\rho_i = 0$ for all $i \in \mathcal{A}(x^*)$ and therefore $\rho = 0$. Therefore, (9.8) reduces to

$$0 = \sum_{j=1}^{p} \mu_j \nabla h_j(x^*),$$

and, since the vectors $\nabla h_j(x^*)$ are linearly independent, by (MFCQ) we conclude that $\mu = 0$. Then $(\rho, \mu) = 0$ and this is a contradiction to $\|(\rho, \mu)\| = 1$. ■

Theorem 9.16 *Let x^* be a strict local minimum of (1.5), which fulfills (MFCQ). Then there exists $\bar{\alpha} > 0$ such that x^* is a local minimum of $P_2(\cdot; \alpha)$ for all $\alpha \geq \bar{\alpha}$. In other words, P_2 is exact at x^*.*

Proof. Since x^* is a strict local minimum of (1.5), there exists $\varepsilon > 0$ such that

$$f(x^*) < f(x) \quad \forall\, x \in B(x^*; \varepsilon) \cap X \setminus \{x^*\}. \tag{9.9}$$

Assume that the statement is not true. Then there exists $(\alpha_k)_{k \geq 0}$ with $\alpha_k \to +\infty$ as $k \to +\infty$ such that, for all $k \geq 0$, x^* is not a local minimum of $P_2(\cdot; \alpha_k)$. On the other hand, for all $k \geq 0$, let x^k be a global minimum of

$$\begin{aligned} &\min\ P_2(x; \alpha_k), \\ &\text{s.t. } x \in \overline{B(x^*; \varepsilon/2)} \end{aligned}$$

where $\overline{B(x^*; \varepsilon/2)}$ denotes the closure of $B(x^*; \varepsilon/2)$. It holds $P_2(x^k; \alpha_k) < P_2(x^*; \alpha_k)$. Indeed, assuming that $P_2(x^k; \alpha_k) = P_2(x^*; \alpha_k)$ would mean that $P_2(x^*; \alpha_k) \leq P_2(x; \alpha_k)$ for all $x \in B(x^*; \varepsilon/2)$ – a contradiction to x^* is not a local minimum of $P_2(\cdot; \alpha_k)$. The following holds for all $k \geq 0$

$$\begin{aligned} f(x^k) + \alpha_k \left\| (g_+(x^k), h(x^k)) \right\|_2 &= P_2(x^k; \alpha_k) \\ &< P_2(x^*; \alpha_k) \\ &= f(x^*) + \alpha_k \left\| (g_+(x^*), h(x^*)) \right\|_2 = f(x^*), \end{aligned} \tag{9.10}$$

where the last equality holds since $\|(g_+(x^*), h(x^*))\|_2 = 0$.

Next, note that since $(x^k)_{k \geq 0}$ lies in the compact set $\overline{B(x^*; \varepsilon/2)}$, there exists a convergent subsequence $(x^{k_l})_{l \geq 0}$ with

$$\lim_{l \to +\infty} x^{k_l} := \overline{x} \in \overline{B(x^*; \varepsilon/2)} \subseteq B(x^*; \varepsilon).$$

By (9.10), we have that for all $l \geq 0$

$$f(x^{k_l}) + \alpha_{k_l} \left\| (g_+(x^{k_l}), h(x^{k_l})) \right\|_2 \leq f(x^*).$$

Taking the limit as $l \to +\infty$, we see that it must hold

$$\|(g_+(\overline{x}), h(\overline{x}))\|_2 = 0,$$

and thus $\overline{x} \in X$. Furthermore,

$$f(\overline{x}) \leq f(x^*)$$

and therefore $\overline{x} = x^*$, because $\overline{x} \in B(x^*; \varepsilon) \cap X$ and x^* is the unique minimum of f on this set.

Then,

$$\lim_{l \to \infty} x^{k_l} = x^*,$$

and thus there exists $l_0 \geq 0$ such that

$$x^{k_l} \in B(x^*; \varepsilon/2) \quad \forall l \geq l_0.$$

By the definition of the sequence $(x^k)_{k \geq 0}$, we have for all $l \geq l_0$

$$P_2(x^{k_l}; \alpha_{k_l}) \leq P_2(x; \alpha_{k_l}) \quad \forall x \in B(x^*; \varepsilon/2).$$

From here we conclude that

$$\nabla P_2(x^{k_l}; \alpha_{k_l}) = 0 \quad \forall l \geq l_0.$$

For all $l \geq l_0$, since $x^{k_l} \in B(x^*; \varepsilon/2)$ and $f(x^{k_l}) \leq f(x^*)$, it must hold $x^{k_l} \notin X$. Therefore, for all $l \geq l_0$,

$$\nabla f(x^{k_l}) + \alpha_{k_l} \nabla r_2(x^{k_l}) = 0. \tag{9.11}$$

From Lemma 9.15, we know that there exists $c > 0$ such that for all $l \geq l_0$

$$\left\|\nabla r_2(x^{k_l})\right\| \geq c.$$

This implies

$$\lim_{l \to +\infty} \alpha_{k_l} \left\|\nabla r_2(x^{k_l})\right\| = +\infty.$$

On the other hand, the sequence $(\nabla f(x^{k_l}))_{l \geq 0}$ is bounded, since $(x^{k_l})_{l \geq 0}$ is bounded and ∇f is continuous. This leads to a contradiction to (9.11) and proves therefore the theorem. ■

Remark 9.17 If x^* is a strict local minimum of (1.5), for which (MFCQ) is fulfilled, then P_q is exact at x^* for all $q \in [1, \infty]$.

10 Sequential Quadratic Programming (SQP) methods

Sequential Quadratic Programming (SQP) methods are among the most important numerical methods for solving of differentiable nonlinear optimization problems.

10.1 Lagrange-Newton iteration

First we will consider optimization problems with equality constraints of type (9.1)

$$\begin{aligned} &\min\ f(x), \\ &\text{such that } h_j(x) = 0, j = 1, \dots, p \\ &x \in \mathbb{R}^n \end{aligned}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $h = (h_1, ..., h_p) : \mathbb{R}^n \to \mathbb{R}^p$ are twice continuously differentiable functions. Let

$$L(x, \mu) = f(x) + \mu^T h(x)$$

be the corresponding Lagrange function and

$$\begin{cases} \nabla_x L(x, \mu) = \nabla f(x) + \nabla h(x)^T \mu = 0 \\ h(x) = 0 \end{cases}$$

the corresponding KKT system of optimality conditions. We would like to use the Newton method for nonlinear equations to solve this system. To this end, we define $\phi : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}^n \times \mathbb{R}^p$ as

$$\phi : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}^n \times \mathbb{R}^p, \quad \phi(x, \mu) = \begin{pmatrix} \nabla f(x) + \nabla h(x)^T \mu \\ h(x) \end{pmatrix},$$

and write the KKT system equivalently as

$$\phi(x, \mu) = 0. \tag{10.1}$$

The so-called Lagrange-Newton iteration for solving (10.1) is as follows:

$$(x^{k+1}, \mu^{k+1}) := (x^k, \mu^k) - \nabla\phi(x^k, \mu^k)^{-1}\phi(x^k, \mu^k) \quad \forall k \geq 0.$$

Algorithm 10.1 (Lagrange-Newton iteration)

1: Choose $(x^0, \mu^0) \in \mathbb{R}^n \times \mathbb{R}^p$, $\varepsilon \geq 0$ and set $k := 0$.

2: If $\left\|\phi(x^k, \mu^k)\right\| \leq \varepsilon$: STOP.

3: Find $(\Delta x^k, \Delta\mu^k) \in \mathbb{R}^n \times \mathbb{R}^p$ as a solution of the linear system of equations

$$\nabla\phi(x^k, \mu^k) \begin{pmatrix} \Delta x^k \\ \Delta\mu^k \end{pmatrix} = -\phi(x^k, \mu^k).$$

4: Set $(x^{k+1}, \mu^{k+1}) := (x^k, \mu^k) + (\Delta x^k, \Delta\mu^k)$, $k := k + 1$ and go to Step 2.

For the convergence analysis, we set $\varepsilon = 0$ and assume that Algorithm 10.1 does not terminate after finitely many iterations.

We know from Theorem 8.7 that Algorithm 10.1 converges superlinearly to an element $(x^*, \mu^*) \in \mathbb{R}^n \times \mathbb{R}^p$ such that $\phi(x^*, \mu^*) = 0$ if $\nabla\phi(x^*, \mu^*)$ is regular. When is this the case? Can we formulate a condition that guarantees this in terms of f and h?

Theorem 10.2 *Let $(x^*, \mu^*) \in \mathbb{R}^n \times \mathbb{R}^p$ be a KKT point of (9.1) such that the following conditions are satisfied:*

(i) (LICQ) holds: the vectors $\nabla h_1(x^*), ..., \nabla h_p(x^*)$ are linearly independent;

(ii) for all $d \in T_2(x^*) = T_{lin}(x^*) = \{d' \in \mathbb{R}^n \mid \nabla h_j(x^*)^T d = 0 \ \forall j = 1, ..., p\} \setminus \{0\}$, it holds

$$d^T \nabla_{xx} L(x^*, \mu^*) d > 0,$$

in other words, $\nabla_{xx} L(x^*, \mu^*)$ is positively definite on $T_2(x^*)$.

Then $\nabla \phi(x^*, \mu^*)$ is regular.

Proof. We will show that the only solution of the linear equation $\nabla \phi(x^*, \mu^*) q = 0$ is $q = 0$. It holds

$$\nabla \phi(x^*, \mu^*) = \begin{pmatrix} \nabla^2_{xx} L(x^*, \mu^*) & \nabla h(x^*)^T \\ \nabla h(x^*) & 0 \end{pmatrix} \in \mathbb{R}^{n+p} \times \mathbb{R}^{n+p},$$

and, for $q = (q^1, q^2)^T \in \mathbb{R}^n \times \mathbb{R}^p$,

$$\nabla \phi(x^*, \mu^*) \begin{pmatrix} q^1 \\ q^2 \end{pmatrix} = 0$$

is equivalent to the linear system of equations

$$\begin{cases} \nabla^2_{xx} L(x^*, \mu^*) q^1 + \nabla h(x^*)^T q^2 = 0 \\ \nabla h(x^*) q^1 = 0 \end{cases}$$

and further to

$$\begin{cases} \nabla^2_{xx} L(x^*, \mu^*) q^1 + \sum_{j=1}^p q_j^2 \nabla h_j(x^*) = 0 \\ \nabla h_j(x^*)^T q^1 = 0, \quad j = 1, \ldots, p. \end{cases} \tag{10.2}$$

Multiplying the first equation on the left by $(q^1)^T$ gives

$$(q^1)^T \nabla^2_{xx} L(x^*, \mu^*) q^1 + \sum_{j=1}^p q_j^2 (q^1)^T \nabla h_j(x^*) = 0.$$

Since

$$(q^1)^T \nabla h_j(x^*) = 0, \quad j = 1, \ldots, p$$

this reduces to

$$(q^1)^T \nabla^2_{xx} L(x^*, \mu^*) q^1 = 0.$$

Furthermore, the second equation in (10.2) also implies that $(q^1) \in T_2(x^*) = T_{lin}(x^*)$. By assumption (ii), we conclude

$$q^1 = 0.$$

Using again the first equation in (10.2), we have

$$\sum_{j=1}^{p} q_j^2 \nabla h_j(x^*) = 0.$$

Assumption (i) gives

$$q_j^2 = 0, \quad j = 1, ..., p,$$

therefore $q = 0$. This proves that $\nabla\phi(x^*, \mu^*)$ is regular. ■

To get a convergent sequence to a KKT point (x^*, μ^*) of (9.1), we also have to choose (x^0, μ^0) in a suitable (unknown) neighbourhood of it.

In the following, we study the optimization problem (1.5)

$$\begin{aligned} &\min\ f(x), \\ &\text{such that } g_i(x) \le 0, i = 1, \dots, m \\ &\qquad\quad h_j(x) = 0, i = 1, \dots, p \\ &\qquad\quad x \in \mathbb{R}^n \end{aligned}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $g = (g_1, \dots, g_m) : \mathbb{R}^n \to \mathbb{R}^m$ and $h = (h_1, \dots, h_p) : \mathbb{R}^n \to \mathbb{R}^p$ are twice continuously differentiable functions. Recall that the KKT system of optimality conditions for (1.5) reads

$$\begin{cases} \nabla_x L(x, \lambda, \mu) = 0 \\ h(x) = 0 \\ \lambda_i \ge 0, g_i(x) \le 0, \lambda_i g_i(x) = 0, i = 1, \dots, m, \end{cases} \tag{10.3}$$

where

$$L(x, \lambda, \mu) = f(x) + \lambda^T g(x) + \mu^T h(x)$$

is the corresponding Lagrange function.

Next, we rewrite (10.3) as a nonlinear equation, as we did for the optimization problem with only equality constraints.

Definition 10.3 (NCP function) A function $\varphi : \mathbb{R}^2 \to \mathbb{R}$ with the property

$$\varphi(a, b) = 0 \quad \Leftrightarrow \quad a \ge 0,\ b \ge 0,\ ab = 0$$

is called NCP (Nonlinear Complementarity Problem) function.

Example 10.4 Examples of NCP functions are:

(i) the minimum function: $\varphi(a, b) = \min\{a, b\}$;

(ii) the Fischer-Burmeister function: $\varphi(a, b) = \sqrt{a^2 + b^2} - a - b$.

Note that these functions are not everywhere differentiable.

Let $\varphi : \mathbb{R}^2 \to \mathbb{R}$ be a NCP function. Then (10.3) is equivalent to

$$\Phi(x, \lambda, \mu) = 0, \tag{10.4}$$

where

$$\Phi : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$$

is given by

$$\Phi(x, \lambda, \mu) = \begin{pmatrix} \nabla_x L(x, \lambda, \mu) \\ \varphi(-g_1(x), \lambda_1) \\ \vdots \\ \varphi(-g_m(x), \lambda_m) \\ h(x) \end{pmatrix}.$$

We would like to use the Newton method to solve (10.4). To this end, we have to guarantee that

(i) the function Φ is continuously differentiable on a neighbourhood of (x^*, λ^*, μ^*);

(ii) $\nabla\Phi(x^*, \lambda^*, \mu^*)$ is regular.

For the particular instance of the minimum function, in order to guarantee (i) we need to impose the additional condition of $g_i(x^*) + \lambda_i^* \neq 0$ for all $i = 1, \ldots, m$. Indeed, the minimum function $\varphi(a, b) = \min\{a, b\}$ is differentiable at every point (a, b) such that $a \neq b$. This will then guarantee that $g_i(x) + \lambda_i \neq 0$ for all $i = 1, \ldots, m$ in a neighbourhood of (x^*, λ^*, μ^*), and further that Φ is differentiable in a neighbourhood of (x^*, λ^*, μ^*).

10.2 The local SQP algorithm

In order to provide a motivation for the local SQP algorithm, we consider again the optimization problem with only equality constraints (9.1). In Algorithm 10.1, for all $k \geq 0$, we set $x^{k+1} := x^k + \Delta x^k$ and $\mu^{k+1} := \mu^k + \Delta\mu^k$, where $(\Delta x^k, \Delta\mu^k)$ is a solution of the linear system of equations

$$\nabla\phi(x^k, \mu^k) \begin{pmatrix} \Delta x^k \\ \Delta\mu^k \end{pmatrix} = -\phi(x^k, \mu^k).$$

This system is equivalent to

$$\begin{cases} \nabla^2_{xx} L(x^k, \mu^k)\Delta x^k + \nabla h(x^k)^T \Delta\mu^k = -\nabla_x L(x^k, \mu^k) \\ \nabla h_j(x^k)^T \Delta x^k = -h_j(x^k), j = 1, \ldots, p. \end{cases} \tag{10.5}$$

By setting

$$H_k := \nabla^2_{xx} L(x^k, \mu^k) \quad \text{and} \quad \mu^k_+ := \Delta\mu^k + \mu^k,$$

(10.5) can be rewritten as

$$\begin{cases} H_k \Delta x^k + \nabla f(x^k) + \nabla h(x^k)^T \mu_+^k = 0 \\ \nabla h_j(x^k)^T \Delta x^k + h_j(x^k) = 0, \ \ j = 1, \ldots, p. \end{cases} \tag{10.6}$$

It is easy to notice that (10.6) is nothing else than the KKT system of the following optimization problem:

$$\begin{aligned} \min \ & \frac{1}{2} \Delta x^T H_k \Delta x + \nabla f(x^k)^T \Delta x. \\ \text{such that } & \nabla h_j(x^k)^T \Delta x + h_j(x^k) = 0, j = 1, \ldots, p \\ & \Delta x \in \mathbb{R}^n. \end{aligned} \tag{10.7}$$

In order to design a numerical method for problems of type (1.5), inspired by the Lagrange-Newton iteration, one could use for the update rule a KKT point of the following quadratic optimization problem

$$\begin{aligned} \min \ & \frac{1}{2} \Delta x^T H_k \Delta x + \nabla f(x^k)^T \Delta x. \\ \text{such that } & \nabla g_i(x^k)^T \Delta x + g_i(x^k) \leq 0, i = 1, \ldots, m \\ & \nabla h_j(x^k)^T \Delta x + h_j(x^k) = 0, j = 1, \ldots, p \\ & \Delta x \in \mathbb{R}^n \end{aligned} \tag{10.8}$$

At this stage, it remains a conjecture. However, we will later establish that it is indeed a reasonable assumption. Building on this, we arrive at the following algorithm.

Algorithm 10.5 (local SQP algorithm)

1: Choose $(x^0, \lambda^0, \mu^0) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$ and set $k := 0$.

2: If (x^k, λ^k, μ^k) is a KKT point of (1.5): STOP.

3: Find a KKT point $(x^{k+1}, \lambda^{k+1}, \mu^{k+1})$ of the quadratic problem

$$\begin{aligned} \min \ & \frac{1}{2}(x - x^k)^T \nabla_{xx}^2 L(x^k, \lambda^k, \mu^k)(x - x^k) + \nabla f(x^k)^T (x - x^k). \\ \text{such that } & \nabla g_i(x^k)^T (x - x^k) + g_i(x^k) \leq 0, i = 1, \ldots, m \\ & \nabla h_j(x^k)^T (x - x^k) + h_j(x^k) = 0, j = 1, \ldots, p. \end{aligned} \tag{10.9}$$

If (10.9) has more than one KKT point, choose $(x^{k+1}, \lambda^{k+1}, \mu^{k+1})$ as the KKT point with the property that

$$\left\| (x^{k+1}, \lambda^{k+1}, \mu^{k+1}) - (x^k, \lambda^k, \mu^k) \right\|$$

is minimal.

4: Set $k := k + 1$ and go to Step 2.

Remark 10.6 If the optimization problem (10.9) has a local minimun, then, according to Corollary 2.3, it has a KKT point.

Theorem 10.7 (local convergence theorem of the SQP algorithm) *Let $(x^*, \lambda^*, \mu^*) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$ be a KKT point of (1.5) such that:*

(i) The strict complementarity condition holds: $g_i(x^*) + \lambda_i^* \neq 0$ for all $i = 1, \dots, m$;

(ii) (LICQ) holds: the vectors $\{\nabla g_i(x^*)\}_{i \in \mathcal{A}(x^*)} \cup \{\nabla h_j(x^*)\}_{j=1}^p$ are linearly independent;

(iii) For all $d \neq 0$ such that $\nabla g_i(x^*)^T d = 0$ for all $i \in \mathcal{A}(x^*)$, and $\nabla h_j(x^*)^T d = 0$ for all $j = 1, \dots, p$, it holds

$$d^T \nabla^2_{xx} L(x^*, \lambda^*, \mu^*) d > 0,$$

in other words, $\nabla^2_{xx} L(x^*, \lambda^*, \mu^*)$ is positively definite on $T_2(x^*)$ (notice that, if (i) holds, then $\mathcal{A}(x^*) = \mathcal{A}_>(x^*)$).

Then there exists $\varepsilon > 0$ such that for all $(x^0, \lambda^0, \mu^0) \in B((x^*, \lambda^*, \mu^*); \varepsilon)$ and every sequence $((x^k, \lambda^k, \mu^k))_{k \geq 0}$ generated by Algorithm 10.5, it holds:

(a) The sequence $((x^k, \lambda^k, \mu^k))_{k \geq 0}$ is well-defined, also meaning that (10.9) has KKT points.

(b) It holds $(x^k, \lambda^k, \mu^k) \to (x^*, \lambda^*, \mu^*)$ superlinearly as $k \to +\infty$.

(c) If $\nabla^2 f$, $\nabla^2 g_i, i = 1, ..., m$, and $\nabla^2 h_j, j = 1, ..., p$, are locally Lipschitz continuous at x^*, then the convergence rate is quadratic.

Proof. Let $\varphi : \mathbb{R}^2 \to \mathbb{R}, \varphi(a, b) = \min\{a, b\}$, and

$$\Phi : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p, \quad \Phi(x, \lambda, \mu) = \begin{pmatrix} \nabla_x L(x, \lambda, \mu) \\ \varphi(-g_1(x), \lambda_1) \\ \vdots \\ \varphi(-g_m(x), \lambda_m) \\ h(x) \end{pmatrix}.$$

Then (x^*, λ^*, μ^*) is a KKT point of (1.5) if and only if it is a solution of (10.4)

$$\Phi(x, \lambda, \mu) = 0.$$

There exists $\varepsilon_1 > 0$ such that Φ is continuously differentiable on $B((x^*, \lambda^*, \mu^*); \varepsilon_1)$ and $\nabla \Phi(x^*, \lambda^*, \mu^*)$ is regular. Then, by Theorem 8.7, there exists $0 < \varepsilon_2 \leq \varepsilon_1$, such that the Newton algorithm applied to (10.4) generates for any starting point $(x^0, \lambda^0, \mu^0) \in B((x^*, \lambda^*, \mu^*); \varepsilon_2)$ a sequence $((x^k, \lambda^k, \mu^k))_{k \geq 0}$ which has following properties

1. $\|(x^{k+1}, \lambda^{k+1}, \mu^{k+1}) - (x^*, \lambda^*, \mu^*)\| \leq \|(x^k, \lambda^k, \mu^k) - (x^*, \lambda^*, \mu^*)\| \quad \forall k \geq 0$;
2. $(x^k, \lambda^k, \mu^k) \to (x^*, \lambda^*, \mu^*)$ superlinearly as $k \to +\infty$;

3. If $\nabla\Phi$ is locally Lipschitz continuous at (x^*, λ^*, μ^*) (which holds if $\nabla^2 f$, $\nabla^2 g_i, i = 1, ..., m$, and $\nabla^2 h_j, j = 1, ..., p$, are locally Lipschitz continuous at x^*), then the convergence rate is quadratic.

We show that if (x^0, λ^0, μ^0) is chosen close enough to (x^*, λ^*, μ^*), then the sequence generated by the Newton algorithm applied to (10.4) coincides with the sequence generated by the local SQP algorithm. More precisely, we will show that, for all $k \geq 0$, $(x^{k+1}, \lambda^{k+1}, \mu^{k+1})$ in the Newton method is nothing other than the element provided by Step 3 of Algorithm 10.5.

A. $(x^{k+1}, \lambda^{k+1}, \mu^{k+1})$ is a KKT point of (10.9)

We have that $(\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu})$ is a KKT point of (10.9) if and only if

$$\begin{cases} \nabla^2_{xx} L(x^k, \lambda^k, \mu^k)(\widetilde{x} - x^k) + \nabla f(x^k) + \sum_{i=1}^m \widetilde{\lambda}_i \nabla g_i(x^k) + \sum_{j=1}^p \widetilde{\mu}_j \nabla h_j(x^k) = 0 \\ \min\left\{-\nabla g_i(x^k)^T(\widetilde{x} - x^k) - g_i(x^k), \widetilde{\lambda}_i\right\} = 0, i = 1, \ldots, m \\ \nabla h_j(x^k)^T(\widetilde{x} - x^k) + h_j(x^k) = 0, j = 1, \ldots, p. \end{cases} \tag{10.10}$$

According to the strict complementarity condition, it holds

$$\begin{aligned} \forall i \in \mathcal{A}(x^*): &\quad g_i(x^*) = 0, \lambda_i^* > 0, \text{ therefore, } -g_i(x^*) - \lambda_i^* < 0 \\ \forall i \notin \mathcal{A}(x^*): &\quad g_i(x^*) < 0, \lambda_i^* = 0, \text{ therefore, } -g_i(x^*) - \lambda_i^* > 0. \end{aligned}$$

We have

$$\begin{aligned} \forall i \in \mathcal{A}(x^*): &\quad \lim_{(x,\lambda,\mu)\to(x^*,\lambda^*,\mu^*)} -g_i(x) - \lambda_i = -g_i(x^*) - \lambda_i^* < 0 \\ \forall i \notin \mathcal{A}(x^*): &\quad \lim_{(x,\lambda,\mu)\to(x^*,\lambda^*,\mu^*)} -g_i(x) - \lambda_i = -g_i(x^*) - \lambda_i^* > 0, \end{aligned}$$

consequently, there exists $\varepsilon_3' > 0$ such that for all $(\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}) \in B\left((x^*, \lambda^*, \mu^*); \varepsilon_3'\right)$ it holds

$$\forall i \in \mathcal{A}(x^*): \quad -g_i(\widetilde{x}) < \widetilde{\lambda}_i \tag{10.11}$$

$$\forall i \notin \mathcal{A}(x^*): \quad -g_i(\widetilde{x}) > \widetilde{\lambda}_i. \tag{10.12}$$

We also have

$$\begin{aligned} \forall i \in \mathcal{A}(x^*): &\quad \lim_{\substack{(x',\lambda',\mu')\to(x^*,\lambda^*,\mu^*) \\ (x,\lambda,\mu)\to(x^*,\lambda^*,\mu^*)}} -g_i(x') - \nabla g_i(x')^T(x - x') - \lambda_i = -g_i(x^*) - \lambda_i^* < 0 \\ \forall i \notin \mathcal{A}(x^*): &\quad \lim_{\substack{(x',\lambda',\mu')\to(x^*,\lambda^*,\mu^*) \\ (x,\lambda,\mu)\to(x^*,\lambda^*,\mu^*)}} -g_i(x') - \nabla g_i(x')^T(x - x') - \lambda_i = -g_i(x^*) - \lambda_i^* > 0, \end{aligned}$$

consequently, there exists $\varepsilon_3'' > 0$ such that for all $(\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}), (x, \lambda, \mu) \in B\left((x^*, \lambda^*, \mu^*); \varepsilon_3''\right)$ it holds

$$\forall i \in \mathcal{A}(x^*): \quad -g_i(\widetilde{x}) - \nabla g_i(\widetilde{x})^T(x - \widetilde{x}) < \lambda_i \tag{10.13}$$

$$\forall i \notin \mathcal{A}(x^*): \quad -g_i(\widetilde{x}) - \nabla g_i(\widetilde{x})^T(x - \widetilde{x}) > \lambda_i. \tag{10.14}$$

Let $\varepsilon_3 := \min\{\varepsilon_3', \frac{\varepsilon_3''}{3}\}$. Then

$$\forall (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}) \in B\left((x^*, \lambda^*, \mu^*); \varepsilon_3\right): \quad \text{(10.11) and (10.12) hold}$$
$$\forall (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}), (x, \lambda, \mu) \in B\left((x^*, \lambda^*, \mu^*); 3\varepsilon_3\right): \quad \text{(10.13) and (10.14) hold.}$$

Let $\varepsilon := \min\{\varepsilon_2, \varepsilon_3\}$, and $(x^0, \lambda^0, \mu^0) \in B\left((x^*, \lambda^*, \mu^*); \varepsilon\right)$. Then, according to property 1., $(x^k, \lambda^k, \mu^k) \in B\left((x^*, \lambda^*, \mu^*); \varepsilon\right)$ for all $k \geq 0$. From (10.13) and (10.14), we obtain that for all $k \geq 0$

$$\forall i \in \mathcal{A}(x^*): \quad -g_i(x^k) - \nabla g_i(x^k)^T (x^{k+1} - x^k) < \lambda_i^{k+1} \tag{10.15}$$
$$\forall i \notin \mathcal{A}(x^*): \quad -g_i(x^k) - \nabla g_i(x^k)^T (x^{k+1} - x^k) > \lambda_i^{k+1}, \tag{10.16}$$

respectively.

The Newton equation reads for all $k \geq 0$

$$\nabla \Phi(x^k, \lambda^k, \mu^k) \begin{pmatrix} x^{k+1} - x^k \\ \lambda^{k+1} - \lambda^k \\ \mu^{k+1} - \mu^k \end{pmatrix} = -\Phi(x^k, \lambda^k, \mu^k). \tag{10.17}$$

We have

$$\nabla \Phi(x, \lambda, \mu) = \begin{pmatrix} \nabla_{xx} L(x, \lambda, \mu) & \nabla g(x)^T & \nabla h(x)^T \\ \varphi_a'(-g_1(x), \lambda_1)(-g_1)_{x_1}(x) \dots \varphi_a'(-g_1(x), \lambda_1)(-g_1)_{x_n}(x) & \varphi_b'(-g_1(x), \lambda_1) \dots 0 & 0 \\ \vdots \quad\quad\quad\quad \vdots & \vdots \quad\quad \vdots & \vdots \\ \varphi_a'(-g_m(x), \lambda_m)(-g_m)_{x_1}(x) \dots \varphi_a'(-g_m(x), \lambda_m)(-g_m)_{x_n}(x) & 0 \dots \varphi_b'(-g_m(x), \lambda_m) & 0 \\ \nabla h(x) & 0 & 0 \end{pmatrix}.$$

We notice that

$$\nabla \varphi(a, b) = \begin{cases} (1, 0)^T, & \text{if } a < b, \\ (0, 1)^T, & \text{if } a > b. \end{cases}$$

This means that (10.17) is for all $k \geq 0$ equivalent to

$$\begin{cases} \nabla_{xx} L(x^k, \lambda^k, \mu^k)(x^{k+1} - x^k) + \sum_{i=1}^m (\lambda_i^{k+1} - \lambda_i^k) \nabla g_i(x^k) + \sum_{j=1}^p (\mu_j^{k+1} - \mu_j^k) \nabla h_j(x^k) = -\nabla L(x^k, \lambda^k, \mu^k) \\ \forall j = 1, ..., p: \nabla h_j(x^k)^T (x^{k+1} - x^k) = -h_j(x^k) \\ \forall i \in \mathcal{A}(x^*): \quad -\nabla g_i(x^k)^T (x^{k+1} - x^k) = -\min\{-g_i(x^k), \lambda_i^k\} = g_i(x^k) \quad (\text{since } -g_i(x^k) < \lambda_i^k) \\ \forall i \notin \mathcal{A}(x^*): \quad \lambda_i^{k+1} - \lambda_i^k = -\min\{-g_i(x^k), \lambda_i^k\} = -\lambda_i^k \quad (\text{since } -g_i(x^k) > \lambda_i^k). \end{cases}$$

Taking into account (10.15) and (10.16), this system is further for all $k \geq 0$ equivalent to

$$\begin{cases} \nabla_{xx} L(x^k, \lambda^k, \mu^k)(x^{k+1} - x^k) + \sum_{i=1}^m (\lambda_i^{k+1} - \lambda_i^k) \nabla g_i(x^k) + \sum_{j=1}^p (\mu_j^{k+1} - \mu_j^k) \nabla h_j(x^k) = -\nabla L(x^k, \lambda^k, \mu^k) \\ \forall i \in \mathcal{A}(x^*): \quad -\nabla g_i(x^k)^T (x^{k+1} - x^k) = -\min\{-g_i(x^k), \lambda_i^k\} = g_i(x^k) \quad (\text{since } -g_i(x^k) < \lambda_i^k) \\ \forall i \notin \mathcal{A}(x^*): \quad \lambda_i^{k+1} - \lambda_i^k = -\min\{-g_i(x^k), \lambda_i^k\} = -\lambda_i^k \quad (\text{since } -g_i(x^k) > \lambda_i^k) \\ \forall j = 1, ..., p: \quad \nabla h_j(x^k)^T (x^{k+1} - x^k) + h_j(x^k) = 0. \end{cases} \tag{10.18}$$

Making use of (10.15) and (10.16), we see that (10.18) is equivalent to

$$\begin{cases} \nabla_{xx} L(x^k, \lambda^k, \mu^k)(x^{k+1} - x^k) + \nabla f(x^k) + \sum_{i=1}^m \lambda_i^{k+1} \nabla g_i(x^k) + \sum_{j=1}^p \mu_j^{k+1} \nabla h_j(x^k) = 0 \\ \forall i \in \mathcal{A}(x^*): \ \min\{-g_i(x^k) - \nabla g_i(x^k)^T (x^{k+1} - x^k), \lambda_i^{k+1}\} = -g_i(x^k) - \nabla g_i(x^k)^T (x^{k+1} - x^k) = 0 \\ \forall i \notin \mathcal{A}(x^*): \ \min\{-g_i(x^k) - \nabla g_i(x^k)^T (x^{k+1} - x^k), \lambda_i^{k+1}\} = \lambda_i^{k+1} = 0 \\ \forall j = 1, ..., p : \nabla h_j(x^k)^T (x^{k+1} - x^k) + h_j(x^k) = 0. \end{cases}$$

This proves that $(x^{k+1}, \lambda^{k+1}, \mu^{k+1})$ is a solution of (10.10) or, equivalently, a KKT point of (10.9).

B. $(x^{k+1}, \lambda^{k+1}, \mu^{k+1})$ is the KKT point of (10.9) that is closest to (x^k, λ^k, μ^k)

First, we show that $(x^{k+1}, \lambda^{k+1}, \mu^{k+1})$ is the only KKT point of (10.9) that lies in $B\left((x^*, \lambda^*, \mu^*); 3\varepsilon\right)$. Indeed, assume that $(\widetilde{x}^{k+1}, \widetilde{\lambda}^{k+1}, \widetilde{\mu}^{k+1}) \in B\left((x^*, \lambda^*, \mu^*); 3\varepsilon\right)$ is a KKT point of (10.9). From (10.15) and (10.16) it yields

$$\begin{cases} \forall i \in \mathcal{A}(x^*): & -g_i(x^k) - \nabla g_i(x^k)^T (\widetilde{x}^{k+1} - x^k) < \widetilde{\lambda}_i^{k+1} \\ \forall i \notin \mathcal{A}(x^*): & -g_i(x^k) - \nabla g_i(x^k)^T (\widetilde{x}^{k+1} - x^k) > \widetilde{\lambda}_i^{k+1}. \end{cases}$$

From Part A we see that $(\widetilde{x}^{k+1}, \widetilde{\lambda}^{k+1}, \widetilde{\mu}^{k+1})$ equivalently fulfills

$$\nabla \Phi(x^k, \lambda^k, \mu^k) \begin{pmatrix} \widetilde{x}^{k+1} - x^k \\ \widetilde{\lambda}^{k+1} - \lambda^k \\ \widetilde{\mu}^{k+1} - \mu^k \end{pmatrix} = -\Phi(x^k, \lambda^k, \mu^k).$$

Since $\nabla \Phi(x^k, \lambda^k, \mu^k)$ is regular and $(x^{k+1}, \lambda^{k+1}, \mu^{k+1})$ is the unique solution of (10.17), it yields $(\widetilde{x}^{k+1}, \widetilde{\lambda}^{k+1}, \widetilde{\mu}^{k+1}) = (x^{k+1}, \lambda^{k+1}, \mu^{k+1})$.

Finally, we assume that $(\bar{x}^{k+1}, \bar{\lambda}^{k+1}, \bar{\mu}^{k+1})$ is a KKT point of (10.9) such that

$$\|(\bar{x}^{k+1}, \bar{\lambda}^{k+1}, \bar{\mu}^{k+1}) - (x^k, \lambda^k, \mu^k)\| < \|(x^{k+1}, \lambda^{k+1}, \mu^{k+1}) - (x^k, \lambda^k, \mu^k)\|.$$

It holds

$$\begin{aligned} & \|(\bar{x}^{k+1}, \bar{\lambda}^{k+1}, \bar{\mu}^{k+1}) - (x^*, \lambda^*, \mu^*)\| \\ \leq & \|(\bar{x}^{k+1}, \bar{\lambda}^{k+1}, \bar{\mu}^{k+1}) - (x^k, \lambda^k, \mu^k)\| + \|(x^k, \lambda^k, \mu^k) - (x^*, \lambda^*, \mu^*)\| \\ < & \|(x^{k+1}, \lambda^{k+1}, \mu^{k+1}) - (x^k, \lambda^k, \mu^k)\| + \|(x^k, \lambda^k, \mu^k) - (x^*, \lambda^*, \mu^*)\| \\ \leq & \|(x^{k+1}, \lambda^{k+1}, \mu^{k+1}) - (x^*, \lambda^*, \mu^*)\| + \|(x^k, \lambda^k, \mu^k) - (x^*, \lambda^*, \mu^*)\| + \|(x^k, \lambda^k, \mu^k) - (x^*, \lambda^*, \mu^*)\| \\ < & 3\varepsilon. \end{aligned}$$

From the above it follows that $(\bar{x}^{k+1}, \bar{\lambda}^{k+1}, \bar{\mu}^{k+1})$ cannot be a a KKT point of (10.9). Consequently, $(x^{k+1}, \lambda^{k+1}, \mu^{k+1})$ is the KKT point of (10.9) that is closest to (x^k, λ^k, μ^k). ■

Bibliography

[1] J.F Bonanns, J.C. Gilbert, C. Lemaréchal, C.A. Sagastizábal: *Numerical Optimization*, Springer Berlin Heidelberg, 2006

[2] R.I. Boţ, J. Fadili, D.-K. Nguyen, *The iterates of Nesterov's accelerated algorithm converge in the critical regimes*, arXiv:2510.22715 (2025)

[3] A. Chambolle, C. Dossal: *On the convergence of the iterates of the "Fast Iterative Shrinkage/Thresholding Algorithm"*, Journal of Optimization Theory and Applications 166(3), 968–982, 2016

[4] C. Geiger, C. Kanzow: *Numerische Verfahren zur Lösung unrestringierter Optimierungsaufgaben*, Springer Berlin Heidelberg, 1999

[5] C. Geiger, C. Kanzow: *Theorie und Numerik restringierter Optimierungsaufgaben*, Springer Berlin Heidelberg, 2002

[6] F. Jarre, J. Stoer: *Optimierung*, Springer Berlin Heidelberg, 2003

[7] U. Jang, E.K. Ryu, *Point convergence of Nesterov's Accelerated Gradient Method: an AI-assisted proof*, arXiv:2510.23513 (2025)

[8] Y. Nesterov: *A method of solving a convex programming problem with convergence rate* $\mathcal{O}(\frac{1}{k^2})$, Soviet Mathematics Doklady 27, 372–376, 1983

[9] J. Nocedal, S.J. Wright: *Numerical Optimization*, Springer Series in Operations Research and Financial Engineering, Springer New York, 2006