Nonlinear Optimization

Lecture Notes

Florian Gensheimer

December 31, 2018

Contents

0. Sources

The major part of these lecture notes is based on the book of Stefan and Michael Ulbrich [2]. Some parts are taken from the script of Sven Krumke [1].

Part I.

Problem and Examples

1. Problem

In this lecture, we consider the following type of optimization problem: We want to minimize a continuous **objective function** $f : X \mapsto \mathbb{R}$ over the non-empty **feasible set** $X \subseteq \mathbb{R}^n$. Our notation is as follows:

$$\begin{aligned} \min \quad & f(x) && (1.1)\\ \text{s.t.} \quad & x \in X && (1.2) \end{aligned}$$

In (1.2), "s.t." means "such that". The condition "$x \in X$" is called **constraint** of the optimization problem. An optimization problem as in (1.1)-(1.2) is called **Nonlinear Optimization Problem** or **Nonlinear Program (NLP)**.
If $X = \mathbb{R}^n$, then the constraint (1.2) is redundant and we obtain the **unconstrained Nonlinear Program (unconstrained NLP)**

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

We consider such problems in Part II.
If $X \neq \mathbb{R}^n$, then the constraint (1.2) is relevant and we speak of a **constrained Nonlinear Program (constrained NLP)**. Usually, X is given by a system of equations and inequalities, i. e.

$$X = \{x \in \mathbb{R}^n : h(x) = 0, g(x) \leq 0\}$$

with continuous functions $h : \mathbb{R}^n \mapsto \mathbb{R}^p$ and $g : \mathbb{R}^n \mapsto \mathbb{R}^m$. The inequality $g(x)$ is meant component-wise, i. e. we define for $y, z \in \mathbb{R}^m$ that

$$y \leq z :\Leftrightarrow y_i \leq z_i \quad \forall i = 1, \ldots, m$$

and

$$y < z :\Leftrightarrow y_i < z_i \quad \forall i = 1, \ldots, m.$$

If f is linear and g, h are affine, i. e. $f(x) = c^\top x$, $g(x) = Ax - b$ and $h(x) = Bx - d$, we obtain a **Linear Optimization Problem (Linear Program)**.
If f is quadratic, i. e.

$$f(x) = \frac{1}{2} x^\top A x + b^\top x + c$$

with $b \in \mathbb{R}^n$, $c \in \mathbb{R}$, $A \in \mathbb{R}^{n \times n}$ and g, h affine, then we get a **Quadratic Optimization Problem**.
If f and all component functions g_i are convex and h is affine, we speak of a **Convex Optimization Problem (Convex Program)**.

2. Examples

As an introduction, we consider three practical examples:

Example 2.1 (Portfolio Optimization)
Our goal is to invest the amount $B > 0$ of money into n stocks for one year such that the expected yield is at least $\rho\%$ and the risk is minimized. We have the following parameters/values:

- The random variable r_i is the yield of stock i at the end of the year
- $x \in \mathbb{R}^n$ with $\sum_{i=1}^n x_i = 1$, $x \geq 0$ describes the composition of the portfolio (we invest the amount $x_i B$ into stock i).
- the yield of the portfolio is then given by

$$R(x) = \frac{\sum_{i=1}^n B(r_i/100)}{B} \cdot 100 = r^\top x.$$

- $\mu \in \mathbb{R}^n$ denotes the expectation value of r.
- $\Sigma \in \mathbb{R}^{n \times n}$ denotes the covariance matrix of r.
- The expected yield is given by $E(R(x)) = \mu^\top x$.
- The variance, which is a measure of the risk of the portfolio, is given by $V(R(x)) = x^\top \Sigma x$.

The Portfolio Optimization Problem can be formulated as follows:

$$\begin{aligned} \min\ & x^\top \Sigma x \\ \text{s.t.}\ & \sum_{i=1}^n x_i = 1 \\ & x \geq 0 \\ & \mu^\top x \geq \rho. \end{aligned}$$

This is a quadratic optimization problem.

Example 2.2 (Cost-Efficient Freight)
We want to send a product with total volume of $1000\,m^3$ per freight. The question, we want to answer, is: Which box sizes (block-shaped, length x_1, depth x_2, height $x_3\,[m]$) shall be used in order to minimize the costs?
We have the following information given:

- The shipping company charges 60 € per sent box and the maximum volume of one box is $1\,m^3$.
- The production of the boxes costs 2 € per m^2 ground and side area and 1 € per m^2 for the caps.
- Only $2000\,m^2$ of cap material is available.
- area of ground and side per box: $x_1x_2 + 2x_1x_3 + 2x_2x_3$.
- area of cap per box: x_1x_2.
- volume per box: $x_1x_2x_3$.
- production costs per box: $f_B(x) = 2(x_1x_2 + 2x_1x_3 + 2x_2x_3) + 1x_1x_2$.
- $n(x)$: number of required boxes $\Rightarrow n(x) = \frac{1000}{x_1x_2x_3}$ (we neglect the integrality condition).
- total costs: $f(x) = f_B(x)n(x) + 60n(x)$.
- constraints: $x_1, x_2, x_3 > 0$, $x_1x_2x_3 \leq 1$, $x_1x_2n(x) \leq 2000$.

By inserting $n(x) = \frac{1000}{x_1x_2x_3}$, we get the following model:

$$\begin{array}{lrl} \min & \dfrac{3000}{x_3} + \dfrac{4000}{x_2} + \dfrac{4000}{x_1} + \dfrac{60000}{x_1x_2x_3} & \\ \text{s.t.} & x_1x_2x_3 & \leq 1 \\ & \dfrac{1000}{x_3} & \leq 2000 \\ & x_1, x_2, x_3 & > 0 \end{array}$$

The inequalities $x_1, x_2, x_3 > 0$ make the model mathematically difficult. Hence, we replace $x_i > 0$ by $x_i \geq l_i$ for some minimum length l_i. The constraint $\frac{1000}{x_3} \leq 2000$ can be replaced by $x_3 \geq \frac{1}{2}$. The drawback of this model is that it is not convex. But the model can be transformed into a convex model. We make the ansatz $x_i = e^{y_i}$. This leads to the following model:

$$\begin{array}{lrl} \min & 3000e^{-y3} + 4000e^{-y_2} + 4000e^{-y_1} + 60000e^{-y_1-y_2-y_3} & \\ \text{s.t.} & e^{y_1+y_2+y_3} & \leq 1 \\ & e^{y_3} & \geq \dfrac{1}{2} \end{array}$$

We apply the natural logarithm ln to the constraints and obtain:

$$y_1 + y_2 + y_3 \leq 0 \quad \text{and} \quad y_3 \geq \ln\left(\frac{1}{2}\right) (= -\ln 2)$$

One can show that applying the natural logarithm to objective function does not destroy the convexity.

Example 2.3 (Optimal Placement of Modules of a Mircoprocessor Chip)
We want to place n modules on a chip in such a way that connected modules are as close to each other as possible.

- For simplicity, we assume that all modules are circles with centres $(x_i, y_i) \in \mathbb{R}^2$ and radius r_i for $i = 1, \ldots, n$.
- The edge set $E \subset \{\{i, j\}; 1 \leq i < j \leq n\}$ describes, which modules are connected.
- We want to ensure that different modules do not overlap.
- The weight $w_{\{i,j\}}$ describes the importance of the connection from module i to j.

We get the following NLP:

$$\begin{aligned} \min_{x,y \in \mathbb{R}^n} & \sum_{\{i,j\} \in E} w_{\{i,j\}} \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \\ \text{s.t.} \quad & (x_i - x_j)^2 + (y_i - y_j)^2 \qquad \geq (r_i + r_j)^2 \quad \forall 1 \leq i < j \leq n \end{aligned}$$

3. Basics and Notation

In this chapter, we recall required basics from analysis and define our notation.

Definition 3.1 Let $x, y \in \mathbb{R}^n$, $M \in \mathbb{R}^{m \times n}$, $\varepsilon > 0$.

a) $\|x\| := \|x\|_2 := \sqrt{x^\top x} = (\sum_{i=1}^n x_i^2)^{\frac{1}{2}}$ is called the **euclidean norm**.

b) $\langle x, y \rangle := \langle x, y \rangle_2 := x^\top y = \sum_{i=1}^n x_i y_i$ is the **euclidean scalar product**.

c) $\|M\| := \max_{\|x\|=1} \|Mx\|$ is the **operator norm** induced by the vector norm $\|\cdot\|$.

d) $B_\varepsilon(x) := \{y \in \mathbb{R}^n : \|x - y\| < \varepsilon\}$ is the **open ε-ball** around x.

e) $\overline{B_\varepsilon(x)} := \{y \in \mathbb{R}^n : \|x - y\| \leq \varepsilon\}$ is the **closed ε-ball** around x.

f) $U \subseteq \mathbb{R}^n$ is a **neighborhood** of $x :\Leftrightarrow B_\varepsilon(x) \subseteq U$ for some $\varepsilon > 0$.

Example 3.2

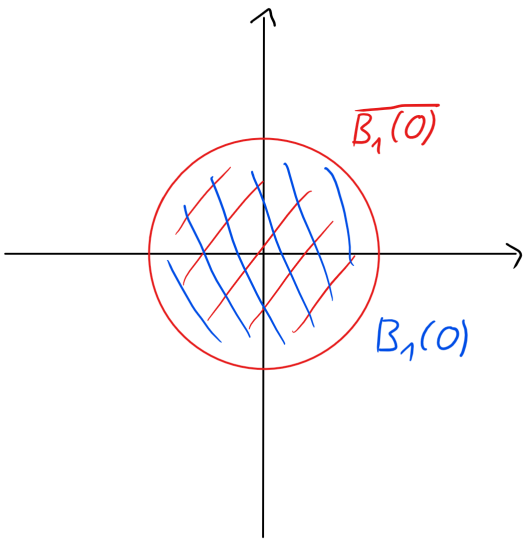


Figure 3.1.: The open and closed unit ball in $\mathbb{R}^2$

Definition 3.3 Let $\Omega \subseteq \mathbb{R}^n$ be open, $f \in C^1(\Omega)$ (i. e. $f : \Omega \mapsto \mathbb{R}$ is continuously differentiable on Ω), $x \in \Omega$.

a) $\nabla f(x) := \begin{pmatrix} \frac{\partial f(x)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{pmatrix}$ is called the **gradient** of f at x.

b) $Df(x) := (\nabla f(x))^\top$

Definition 3.4 Let $\Omega \subseteq \mathbb{R}^n$ be open, $f \in C^2(\Omega)$, $x \in \Omega$.

$$\nabla^2 f(x) := \left(\frac{\partial^2 f(x)}{\partial x_i \partial x_j}\right)_{i,j=1,\dots,n} = \begin{pmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \cdots & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{pmatrix}$$

is called the **Hessian matrix** of f at x.

Definition 3.5 Let $\Omega \subseteq \mathbb{R}^n$ be open, $F \in C^1(\Omega, \mathbb{R}^m)$ (i. e. $F : \Omega \mapsto \mathbb{R}^m$ is continuously differentiable on Ω) and $x \in \Omega$.

$$F'(x) := \left(\frac{\partial F_i}{\partial x_j}(x)\right)_{\substack{i=1,\dots,m \\ j=1,\dots,n}} := \begin{pmatrix} \frac{\partial F_1}{\partial x_1}(x) & \frac{\partial F_1}{\partial x_2}(x) & \cdots & \frac{\partial F_1}{\partial x_n}(x) \\ \frac{\partial F_2}{\partial x_1}(x) & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \frac{\partial F_m}{\partial x_1}(x) & \cdots & \cdots & \frac{\partial F_m}{\partial x_n}(x) \end{pmatrix}$$

is called the **Jacobian matrix** of F at x.
We will also use the notation

$$\nabla F(x) = (\nabla F_1(x), \dots, \nabla F_m(x)) = F'(x)^\top \in \mathbb{R}^{n \times m}.$$

Example 3.6 Consider the quadratic function $f(x) = \frac{1}{2} x^\top A x + b^\top x + c$, with $c \in \mathbb{R}$, $b \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n \times n}$ symmetric. Then, it holds that $\nabla f(x) = Ax + b$ and $\nabla^2 f(x) = A$.

Definition 3.7 Let f and g be two functions on some subset of $\mathbb{R}$.

a) $f(x) \in \mathcal{O}(g(x))$ as $x \mapsto a$ (often: $a = 0$) $:\Leftrightarrow$
there exists some $\delta > 0$ and $M > 0$: $|f(x)| \leq M \cdot |g(x)|$ if $0 < |x - a| < \delta$.

b) Let $g(x) \neq 0$ for all x.
$f(x) \in o(g(x)) :\Leftrightarrow \lim\limits_{x \mapsto 0} \frac{f(x)}{g(x)} = 0$.

Definition 3.8 Let $f : \Omega \mapsto \mathbb{R}$, $\Omega \subseteq \mathbb{R}^n$.

1. $x^* \in \Omega$ is a **local minimum** of $f :\Leftrightarrow$
there exists a neighborhood U of x^* such that for all $x \in \Omega \cap U$:

$$f(x^*) \leq f(x)$$

2. $x^* \in \Omega$ is a **strict local minimum** of $f :\Leftrightarrow$
   there exists a neighborhood U of x^* such that for all $x \in (\Omega \cap U) \setminus \{x^*\}$:

$$f(x^*) < f(x)$$

3. $x^* \in \Omega$ is a **global minimum** of $f :\Leftrightarrow$

$$f(x^*) \leq f(x) \quad \text{for all } x \in \Omega$$

4. $x^* \in \Omega$ is a **strict global minimum** of $f :\Leftrightarrow$

$$f(x^*) < f(x) \quad \text{for all } x \in \Omega \setminus \{x^*\}$$

Remark 3.9 Finding a global minimum algorithmically can be arbitrarily hard. The number of local minima can be very large and every local minimum is a potential candidate for a global minimum. Hence, we will not try to find global minima in this lecture. Instead, we will learn how to characterize and compute local minima.

Theorem 3.10 (Taylor's theorem in $\mathbb{R}$) *Let $g \in C^{p+1}([a,b])$. If $t \in [a,b]$, $\delta > 0$ with $t + \delta \in [a,b]$, then there exists some $\xi \in (t, t+\delta)$ such that*

$$g(t+\delta) = \sum_{k=0}^{p} \frac{g^{(k)}(t)}{k!}\delta^k + R_p(t,\delta),$$

where $R_p(t,\delta) = \frac{g^{(p+1)}(\xi)}{(p+1)!}\delta^{p+1}$.

Application of Taylor/Variant of Taylor:
Let $f : \Omega \mapsto \mathbb{R}$, $\Omega \subseteq \mathbb{R}^n$ open, $f \in C^2(\Omega)$. Let $x \in \Omega$ and $h \in \mathbb{R}^n$.
We apply Theorem 3.10 to $g(t) := f(x + t \cdot h)$ at $t = 0$:
It follows that

$$f(x+h) = f(x) + \nabla f(x)^\top \cdot h + \frac{1}{2} h^\top \nabla^2 f(x+\theta h) \cdot h$$

for some $0 < \theta < 1$.
We can conclude that

$$\begin{aligned} f(x+h) &= f(x) + \nabla f(x)^\top \cdot h + \frac{1}{2}\nabla^2 f(x) h + \frac{1}{2} h^\top (\nabla^2 f(x+\theta h) - \nabla^2 f(x)) \cdot h \\ &= f(x) + \nabla f(x)^\top h + \frac{1}{2} h^\top \nabla^2 f(x) h + o(\|h\|_2^2) \end{aligned}$$

The last equation shows that a smooth function behaves locally like a quadratic function.

Theorem 3.11 (Taylor's Theorem in $\mathbb{R}^n$) *Let $\Omega \subseteq \mathbb{R}^n$ be open, $g : \Omega \mapsto \mathbb{R}$ with $g \in C^2(\Omega)$. Let $x_0 \in \Omega$ and $\delta > 0$ with $\overline{B_\delta(x_0)} \subseteq \Omega$.*
Then there exists some $M = M_\delta > 0$ such that for all h with $\|h\| \leq \delta$, it holds:

$$g(x_0 + h) = g(x_0) + Dg(x_0) \cdot h + r(h)$$

with $\|r(h)\|_\infty \leq M \cdot \|h\|_\infty^2$.

Definition 3.12 Let $\Omega \subseteq \mathbb{R}^n$ be open and $f \in C^1(\Omega)$.

$$x \in \Omega \text{ is called } \textbf{stationary point} \text{ of } f :\Leftrightarrow \nabla f(x) = 0.$$

Definition 3.13 Let $A \in \mathbb{R}^{n \times n}$

a) A is called **positive semidefinite** $:\Leftrightarrow \ \forall x \in \mathbb{R}^n : x^\top A x \geq 0$.

b) A is called **positive definite** $:\Leftrightarrow \ \forall x \in \mathbb{R}^n \setminus \{0\} : x^\top A x > 0$.

If A is symmetric, then A being positive semidefinite (positive definite) is equivalent to the fact that all eigenvalues of A are non-negative (positive).

Part II.

Unconstrained Optimization

4. Introduction

In part II, we consider the theory and numerics of unconstrained nonlinear programs (nonlinear NLPs), i. e. problems of the form

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

with objective function $f : \mathbb{R}^n \mapsto \mathbb{R}$. First, we derive optimality conditions.

5. Optimality Conditions

Goal: Necessary and sufficient conditions for $\bar{x}$ being a (local) minimum.
In this chapter, we assume that $f : U \subseteq \mathbb{R}^n \mapsto \mathbb{R}$ with U open and f differentiable.

Theorem 5.1 (First-Order Necessary Condition) *Let $f \in C^1(U)$. Let $\bar{x} \in U$ be a local minimum of f.*
Then it holds: $\nabla f(\bar{x}) = 0$ (i. e. $\bar{x}$ is a stationary point)

Proof. Let $\phi(t) := f(\bar{x} - t\nabla f(\bar{x}))$.
ϕ is continuously differentiable for small $|t|$.
$\phi'(0) = -Df(\bar{x}) \cdot \nabla f(\bar{x}) = -\|\nabla f(\bar{x})\|_2^2$.
If $\nabla f(\bar{x}) \neq 0$, then $\phi'(0) < 0$.
$\Rightarrow \phi(\tau) < \phi(0)$ for small $\tau > 0$, which is a contradiction to $\bar{x}$ being a local minimum. ∎

Remark 5.2 Theorem 5.1 is necessary, but not sufficient.
Note: $\nabla(-f) = -\nabla f$
Hence, every stationary point of f is a stationary point of $-f$ (cannot distinguish between maximum and minimum).

Definition 5.3 $\bar{x}$ is called **saddle point** of $f :\Leftrightarrow \bar{x}$ is a stationary point of f and $\bar{x}$ is neither a maximum nor a minimum of f.

Example 5.4 $f : \mathbb{R}^2 \mapsto \mathbb{R}$, $f(x) = x_1^2 - x_2^2$
$\nabla f(x) = \begin{pmatrix} 2x_1 \\ -2x_2 \end{pmatrix}$
$\bar{x} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ is a stationary and a saddle point (since f is increasing in x_1-direction and decreasing in x_2-direction).

Theorem 5.5 (Second-Order Necessary Condition) *Let $f \in C^2(U)$ and let $\bar{x} \in U$ be a local minimum. Then it holds:*

a) $\nabla f(\bar{x}) = 0$

b) $\nabla^2 f(\bar{x})$ *is positive semidefinite (i. e.* $\forall d \in \mathbb{R}^n : d^\top \nabla^2 f(\bar{x}) d \geq 0$*)*

Proof. 1. already shown in Theorem 5.1

2. Assume there exists some $h \in \mathbb{R}^n : h^\top \nabla^2 f(\bar{x}) h < 0$

$$\Rightarrow f(\bar{x}+h) = f(\bar{x}) + \underbrace{\nabla f(\bar{x})^\top}_{=0} h + \frac{1}{2} h^\top \nabla^2 f(\bar{x}+\theta h) h = f(\bar{x}) + \frac{1}{2} h^\top \nabla^2 f(\bar{x}+\theta h) h$$

for some $0 < \theta < 1$.
Since $\nabla^2 f$ is continuous, it follows that $\frac{1}{2} h^\top \nabla^2 f(\bar{x}+\theta h) h < 0$ (if θh is not small enough, one could divide h by a sufficiently large number)
$\Rightarrow f(\bar{x}+h) < f(\bar{x})$, which is a contradiction to $\bar{x}$ being a local minimum. ∎

Remark 5.6 Theorem 5.5 is not sufficient:
Consider $f(x) = x^3$ and $\bar{x} = 0$.
Theorem 5.5 a) and b) are satisfied, but $\bar{x}$ is a saddle point.

Theorem 5.7 (Second-Order Sufficient Condition) *Let $f \in C^2(U)$ and let $\bar{x} \in U$ satisfy that*

a) $\bar{x}$ is stationary

b) $\nabla^2 f(\bar{x})$ is positive definite

Then $\bar{x}$ is a (strict) local minimum of f.

Proof. Let a) and b) be satisfied.
Then there exists some $\mu > 0 : \min\limits_{\|d\|=1} d^\top \nabla^2 f(\bar{x}) d = \mu$.
$\Rightarrow d^\top \nabla f(\bar{x}) d \geq \mu \cdot \|d\|^2 \quad \forall d \in \mathbb{R}^n$.
Taylor's theorem and $\bar{x}$ stationary gives:
there exists some $\varepsilon > 0$ such that for $d \in B_\varepsilon(0)$ it holds:

$$f(\bar{x}+d) - f(\bar{x}) = \frac{1}{2} d^\top \nabla^2 f(\bar{x}) d + o(\|d\|^2) \geq \frac{1}{2} \mu \|d\|^2 + o(\|d\|^2) \geq \frac{\mu}{4} \|d\|^2$$

Hence, $\bar{x}$ is a (strict) local minimum. ∎

Remark 5.8 Theorem 5.7 a) and b) are sufficient, but not necessary.
Consider $f(x) = x^4$; $\bar{x} = 0$ is a strict (global) minimum.
$\nabla f(x) = f'(x) = 4x^3$
$\nabla f(\bar{x}) = f'(0) = 0$
$\nabla^2 f(x) = f''(x) = 12x^2$
$\nabla^2 f(\bar{x}) = f''(0) = 0$
Hence, $d^\top \nabla^2 f(0) d = d \cdot 0 \cdot d = 0$ for all $d \in \mathbb{R}$ and $\nabla^2 f(\bar{x})$ is positive semidefinite, but not positive definite.

6. Convex Functions

In this chapter, we consider convex functions. We will see that they have the nice property that every local minimum is a global minimum and the local minima are exactly the stationary points.

Definition 6.1 $X \subseteq \mathbb{R}^n$ is called **convex** $:\Leftrightarrow \forall x, y \in X \; \forall \lambda \in [0,1] : (1-\lambda)x + \lambda y \in X$.

Definition 6.2 Let $f : X \mapsto \mathbb{R}$, $X \subseteq \mathbb{R}^n$ convex

a) f is called **convex** $:\Leftrightarrow \forall x, y \in X \; \forall \lambda \in [0,1]$:

$$f((1-\lambda)x + \lambda y) \leq (1-\lambda)f(x) + \lambda f(y)$$

b) f is called **strictly convex** $:\Leftrightarrow \forall x, y \in X, x \neq y, \; \forall \lambda \in (0,1)$:

$$f((1-\lambda)x + \lambda y) < (1-\lambda)f(x) + \lambda f(y)$$

c) f is called **uniformly convex** $:\Leftrightarrow \exists \mu > 0 : \; \forall x, y \in X \; \forall \lambda \in [0,1]$:

$$f((1-\lambda)x + \lambda y) + \mu\lambda(1-\lambda) \cdot \|y - x\|_2^2 \leq (1-\lambda)f(x) + \lambda f(y)$$

Graphically, the convexity of a function f means that the graph between any two points $(x, f(x))$ and $(y, f(y))$ of f lies below (or on) the connecting line of the two points (see Figure 6.1).

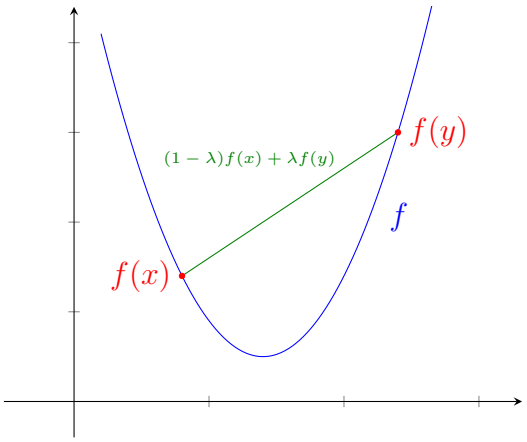


Figure 6.1.: A convex function

For continuously differentiable functions, we obtain the following equivalences:

Theorem 6.3 *Let $f \in C^1(U)$, $U \subseteq X$, U open and X convex. Then it holds:*

a) f is convex $:\Leftrightarrow \forall x, y \in X$:

$$\nabla f(x)^\top (y - x) \leq f(y) - f(x)$$

b) f is strictly convex $:\Leftrightarrow \forall x, y \in X, x \neq y$:

$$\nabla f(x)^\top (y - x) < f(y) - f(x)$$

c) f is uniformly convex $:\Leftrightarrow \exists \mu > 0 \ \forall x, y \in X$:

$$\nabla f(x)^\top (y - x) + \mu \|y - x\|^2 \leq f(y) - f(x)$$

Proof.

1. "$\Rightarrow$:"
   Let f be convex, let $x, y \in X$, $0 < \lambda \leq 1$. Then it holds:

   $$\frac{f(x + \lambda(y - x)) - f(x)}{\lambda} \leq \frac{(1 - \lambda) f(x) + \lambda f(y) - f(x)}{\lambda} = f(y) - f(x)$$

   Now: considering the limit $\lambda \mapsto 0^+$ yields to the right-hand side of a) since

   $$\lim_{\lambda \to 0^+} \frac{f(x + \lambda(y - x)) - f(x)}{\lambda} = \nabla f(x)^\top (y - x)$$

   "$\Leftarrow$:"
   Let $x, y \in X$, $0 \leq \lambda \leq 1$.
   Define $x_\lambda := (1 - \lambda)x + \lambda y$.
   We need to show that $(1 - \lambda) f(x) + \lambda f(y) - f(x_\lambda) \geq 0$.
   It follows that

   $$\begin{aligned} (1 - \lambda) f(x) + \lambda f(y) - f(x_\lambda) &= (1 - \lambda)(f(x) - f(x_\lambda)) + \lambda (f(y) - f(x_\lambda)) \\ &\geq (1 - \lambda) \nabla f(x_\lambda)^\top \cdot (x - x_\lambda) + \lambda \nabla f(x_\lambda)^\top (y - x_\lambda) \\ &= 0 \end{aligned}$$

2. "$\Rightarrow$" Let f be strictly convex.
   For all $x, y \in X$, $x \neq y$ and $z := \frac{x+y}{2}$, it holds:

   $$f(z) < \frac{1}{2}(f(x) + f(y)) \tag{6.1}$$

   $\Rightarrow f(z) - f(x) < \frac{1}{2}(f(y) - f(x))$
   $\overset{a)}{\Rightarrow} f(z) - f(x) \geq \nabla f(x)^\top (z - x) \overset{\text{def. } z}{=} \frac{1}{2} \nabla f(x)^\top (y - x)$
   $\Rightarrow \nabla f(x)^\top (y - x) \leq 2(f(z) - f(x)) \overset{(6.1)}{<} f(y) - f(x)$.
   "$\Leftarrow$:" Similar to "$\Leftarrow$" in part a), just use $>$ insted of $\geq$.

3. "⇒:"

$$\begin{aligned}\nabla f(x)^\top (y-x) &\overset{\text{see a)}}{=} \lim_{\lambda \to 0^+} \frac{f(x_\lambda) - f(x)}{\lambda} \\ &\leq \lim_{\lambda \to 0^+} \frac{(1-\lambda)f(x) + \lambda f(y) - \mu\lambda(1-\lambda)\|y-x\|^2 - f(x)}{\lambda} \\ &= f(y) - f(x) - \mu\|y-x\|^2\end{aligned}$$

"⇐:" We will use that

$$\|x - x_\lambda\| = \lambda\|y-x\| \quad \text{and } \|y - x_\lambda\| = (1-\lambda)\|y-x\| \tag{6.2}$$

■

It follows that

$$\begin{aligned}&(1-\lambda)f(x) + \lambda f(y) - f(x_\lambda) \\ &= (1-\lambda)(f(x) - f(x_\lambda)) + \lambda(f(y) - f(x_\lambda)) \\ &\geq (1-\lambda)(\nabla f(x_\lambda)^\top (x - x_\lambda) + \mu||x-x_\lambda||_2^2) + \lambda(\nabla f(x_\lambda)^\top (y - x_\lambda) + \mu||y-x_\lambda||_2^2) \\ &= \nabla f(x_\lambda)^\top ((1-\lambda)(x-x_\lambda) + \lambda y - x_\lambda) + \mu((1-\lambda)||x-x_\lambda||_2^2 + \lambda||y-x_\lambda||_2^2) \\ &\overset{(6.2)}{=} \mu((1-\lambda)\lambda^2 + \lambda(1-\lambda)^2)||y-x||^2 \\ &= \mu\lambda(1-\lambda)||y-x||_2^2\end{aligned}$$

Graphically, Theorem 6.3 a) says that the tangent of f in an arbitrary point x lies below or on the graph of f (see Figure 6.2).

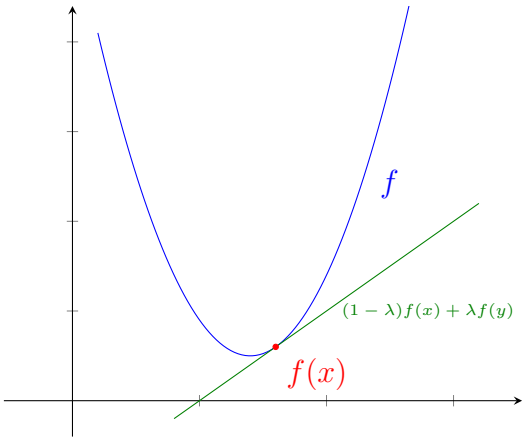


Figure 6.2.: A continuously differentiable convex function

Theorem 6.4 *Let $f \in C^2(X)$, $X \subseteq \mathbb{R}^n$ open and convex. Then it holds:*

a) f is convex $\Leftrightarrow \nabla^2 f(x)$ is positive semidefinite for all $x \in X$

b) f is strictly convex if $\nabla^2 f(x)$ is positive definite for all $x \in X$

*c) f is uniformly convex $\Leftrightarrow \nabla^2 f(x)$ is **uniformly positive definite** for all $x \in X$ (i. e. there exists some $\mu > 0$ auch that $\forall x \in X \ \forall d \in \mathbb{R}^n : d^\top \nabla^2 f(x) d \geq \mu \|d\|^2$)*

Proof.

1. "$\Rightarrow$": Let f be convex, $x \in X$, $d \in \mathbb{R}^n$
   Since X is open, it follows that there exists some $\tau = \tau(x,d) > 0 : \ \forall t \in [0,\tau] : x + t \cdot d \in X$
   For all $0 < t \leq \tau$, it holds:

$$0 \overset{\text{Thm.6.3}}{\leq} f(x+td) - f(x) - t \cdot \nabla f(x)^\top d \overset{\text{Taylor}}{=} \frac{t^2}{2} d^\top \nabla^2 f(x) d + o(t^2)$$

   Multiplying by $\frac{2}{t^2}$ and taking the limit $t \to 0^+$ gives the result.
   "$\Leftarrow$:"
   Taylor's Theorem says that there exists some $\sigma \in [0,1]$ such that

$$f(y) - f(x) = \nabla f(x)^\top (y-x) + \frac{1}{2} \underbrace{(y-x)^\top \nabla^2 f(x + \sigma(x-y))(y-x)}_{\geq 0} \geq \nabla^f(x)^\top (y-x)$$

   With Theorem 6.3, it follows that f is convex.

2. similar to a)

3. "$\Rightarrow$:" As in a), for all $x \in X$ and $d \in \mathbb{R}^n \setminus \{0\}$, there exists some $\tau = \tau(x,d) > 0$ such that $x + td \in X$ for all $t \in [0,\tau]$.
   Similar to a), it follows that

$$0 \leq f(x+td) - f(x) - t \cdot \nabla f(x)^\top d - \mu \cdot \|td\|^2 = \frac{t^2}{2} d^\top \nabla^2 f(x) d - t^2 \mu \cdot \|d\|^2 + o(t^2)$$

   Multiplyiing by $\frac{2}{t^2}$ and taking the limit $t \to 0^+$ gives:

$$d^\top \nabla^2 f(x) d \geq 2\mu \|d\|^2.$$

   "$\Leftarrow$": By Taylor's Theorem there exists some $\sigma \in [0,1]$ such that

$$\begin{aligned} f(y) - f(x) &= \nabla f(x)^\top (y-x) + \frac{1}{2} (y-x)^\top \nabla^2 f(x + \sigma(y-x)) \cdot (y-x) \\ &\overset{\text{Ass.}}{\geq} \nabla f(x)^\top (y-x) + \frac{\mu}{2} \|y-x\|^2 \end{aligned}$$

   From Theorem 6.3, it follows that f is uniformly convex. ∎

Example 6.5

a) $f(x) = a^\top x + b$ with $a \in \mathbb{R}^n$, $b \in \mathbb{R}$ is convex.

b) $f(x) = \max\limits_{1 \le i \le m} \{a_i^\top x + b_i\}$ is convex, but not strictly convex

c) $f(x) = \|x\|$ is convex

d) Let $f(x) = \frac{1}{2} x^\top A x + b^\top x + c$.
Then, it holds:

   i) f is convex $\Leftrightarrow$ A is positive semidefinite

   ii) f is strictly convex $\Leftrightarrow$ f is uniformly convex $\Leftrightarrow$ A is positive definite

e) $f(x) = x^4$ is strictly convex, but not uniformly convex.
$\nabla^2 f(x) = f''(x) = 12x^2$, $\nabla^2 f(0) = 0$
Hence, $\nabla^2 f(x)$ is not positve definite in $x = 0$.
This shows that the condition of Theorem 6.4 b) is only sufficient, but not necessary.

f) $f(x, y) = x^2 + y^2 = \frac{1}{2} \begin{pmatrix} x \\ y \end{pmatrix}^\top \underbrace{\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}}_{=:A} \begin{pmatrix} x \\ y \end{pmatrix}$

Since A is positive definite, it follows from d) that f is strictly convex. This function is illustrated in Figure 6.3.

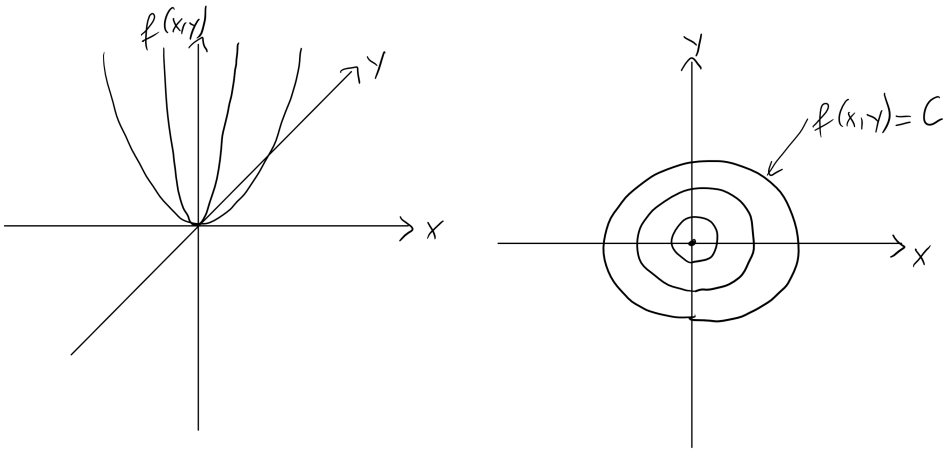


Figure 6.3.: Function $f(x, y) = x^2 + y^2$

Theorem 6.6 *Let $f : X \subseteq \mathbb{R}^n \mapsto \mathbb{R}$, X convex and f convex. Then, it holds:*

a) If $\bar{x}$ is a local minimum of f on X, then $\bar{x}$ is also global minimum.

b) If f is strictly convex, then f has at most one local minimum on X (which is then a strict global minimum)

c) If X is open, $f \in C^1(X)$ and $\bar{x} \in X$ a stationary point, then $\bar{x}$ is a global minimum of f on X.

Proof.

a) Let $\bar{x}$ be a local minimum.
Let us assume that there is a $x \in X$ with $f(x) < f(\bar{x})$.
Then it holds for all $t \in [0, 1]$ that

$$f(\bar{x} + t(x - \bar{x})) \leq (1 - t)f(\bar{x}) + tf(x) < (1 - t)f(\bar{x}) + tf(\bar{x}) = f(\bar{x}), \qquad \blacksquare$$

which is a contradiction to $\bar{x}$ being a local minimum.

b) Assume that $\bar{x}$ and $\bar{y}$ with $\bar{x} \neq \bar{y}$ are local minima on X.
From a), it follows $\bar{x}$ and $\bar{y}$ are global minima
$\Rightarrow \ \forall x \in X : f(x) \geq f(\bar{x}) = f(\bar{y})$
Let $x = \frac{\bar{x}+\bar{y}}{2} \in X$.
$\Rightarrow f(x) \overset{\text{strict convex}}{<} \frac{f(\bar{x})+f(\bar{y})}{2} = f(\bar{x}) \leq f(x)$, which is a contradiction.

c) By Theorem 6.3, it holds for all $x \in X$ that $f(x) - f(\bar{x}) \geq \underbrace{\nabla f(\bar{x})^\top}_{=0}(x - \bar{x}) = 0$

Hence, $\bar{x}$ is a global minimum.

Remark 6.7 Theorem 6.6 allows to identify NLPs for which we can find global minima by finding local minima or even just stationary points.

7. Gradient Method

In this chapter we consider the problem $\min_{x \in \mathbb{R}^n} f(x)$ with $f : \mathbb{R}^n \to \mathbb{R}$ and $f \in C^1(\mathbb{R}^n)$.
The goal in this and the following chapters is to find a stationary point by using descent directions. Such algorithms are called **descent methods**. We will prove and analyze convergence properties for descent methods.
Descent methods have the following general structure:

Algorithm 7.1 General outline of descent methods

1: Choose starting point $x^0 \in \mathbb{R}^n$
2: **for** all $k = 0, 1, 2, \ldots$ **do**
3: **if** $\nabla f(x^k) = 0$ **then**
4: Stop
5: Compute a descent direction $s^k \in \mathbb{R}^n$, i. e. a direction with $\nabla f(x^k)^\top \cdot s^k < 0$
6: Compute a step size $\sigma_k > 0$ such that $f(x^k + \sigma_k s^k) < f(x^k)$ and $f(x^k) - f(x^k + \sigma_k s^k)$ is sufficiently large
7: Set $x^{k+1} = x^k + \sigma_k \cdot s^k$

Definition 7.2 $s \in \mathbb{R}^n \setminus \{0\}$ is called **descent direction** of f in x if

$$\nabla f(x)^\top \cdot s < 0.$$

Remark 7.3

1. The slope of f in direction s is

$$\lim_{t \to 0^+} \frac{f(x + ts) - f(x)}{\|t \cdot s\|} = \frac{\nabla f(x)^\top s}{\|s\|}$$

2. Let $\phi(t) := f(x + t \cdot s)$.
$\phi'(0) = \nabla f(x)^\top \cdot s$, i. e. $\nabla f(x)^\top \cdot s$ is the slope of $\phi(t)$ at $t = 0$.

Example 7.4 Consider $f(x) = -x_1 - x_2^2$, $x \in \mathbb{R}^2$
It follows that $\nabla f(x) = \begin{pmatrix} -1 \\ -2x_2 \end{pmatrix}$ and $\nabla^2 f(x) = \begin{pmatrix} 0 & 0 \\ 0 & -2 \end{pmatrix}$.
Let $s = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ and $x = 0$.
Then it holds that $\phi'(0) = \nabla f(0)^\top \cdot s = \begin{pmatrix} -1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = 0$.

Thus, s is not a descent direction.
But f descreases along s since for $t > 0$, it holds:

$$\phi(t) = f(0 + t \cdot s) = -t^2 < 0 = \phi(0)$$

Hence, for s being a descent direction, it is not sufficient that f decreases along s.
However, for $d = \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} \in \mathbb{R}^2$ with $d_1 > 0$, d is a descent direction since

$$\nabla f(0)^\top \cdot d = \begin{pmatrix} -1 & 0 \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} = -d_1 < 0.$$

Definition 7.5 Let $f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^1(\mathbb{R}^\kappa)$, $x \in \mathbb{R}^n$ with $\nabla f(x) \neq 0$. Let

$$d = \arg\min_{\|d\|=1} \nabla f(x)^\top d. \tag{7.1}$$

Then, any $s := \lambda d$ with $\lambda > 0$ is called **direction of steepest descent** of f in x.

Theorem 7.6 *Problem (7.1) has the unique optimal solution*

$$d = -\frac{\nabla f(x)}{\|\nabla f(x)\|}.$$

(i. e. s is a direction of steepest descent $\Leftrightarrow \exists \lambda > 0 : s = -\lambda \cdot \nabla f(x)$)

Proof. Apply Cauchy-Schwartz Inequality: $|v^\top w| \leq \|v\| \cdot \|w\| \ \forall v, w \in \mathbb{R}$.
The latter is satisfied with equality if and only if v and w are lienarly independent.
For $d \in \mathbb{R}^n$, $\|d\| = 1$, it holds:

$$\nabla f(x)^\top \cdot d \geq -\|\nabla f(x)\| \cdot \|d\| = -\|\nabla f(x)\|$$

and equality holds if and only if $d = -\frac{\nabla f(x)}{\|\nabla f(x)\|}$ ∎

The **Gradient method** uses $s^k = -\nabla f(x^k)$, i. e. the direction of steepest descent in Algorithm 7.1.
An open question is: How do we choose the step length σ_k?
Possibility 1: **Exact line search**
Solve

$$\lambda_k := \arg\min\{f(x^k + \lambda \cdot s^k) : \lambda \geq 0\} \tag{7.2}$$

(7.2) is typically hard to solve in practice.
Possibility 2: **Armijo line search**
Let $\beta \in (0, 1)$ (e. g. $\beta = \frac{1}{2}$) and $\gamma \in (0, 1)$ (e. g. $\gamma = \frac{1}{100}$).
Compute the largest number $\sigma_k \in \{1, \beta, \beta^2, \dots\}$ such that

$$f(x^k + \sigma_k s^k) - f(x^k) \leq \underbrace{\sigma_k \cdot \gamma}_{>0} \cdot \underbrace{\nabla f(x^k)^\top \cdot s^k}_{<0}. \tag{7.3}$$

This line search is easy to implement and always applicable (see Theorem 7.7).

Theorem 7.7 *Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$, $f \in C^1(U)$, U open. Let $\gamma \in (0,1)$. If $x \in U$ and $s \in \mathbb{R}^n$ is a descent direction, then there exists some $\bar{\sigma} > 0$ with*

$$f(x + \sigma s) - f(x) \leq \sigma\gamma \cdot \nabla f(x)^\top \cdot s \quad \forall \sigma \in [0, \bar{\sigma}]$$

Proof. $\sigma = 0 : \checkmark$
Let $\sigma > 0$ be sufficiently small. For $x + \sigma s \in U$, it holds:

$$\frac{f(x+\sigma s) - f(x)}{\sigma} - \gamma \nabla f(x)^\top \cdot s \overset{\sigma \to 0^+}{\longrightarrow} \nabla f(x)^\top s - \gamma \nabla f(x)^\top s = (1-\gamma)\nabla f(x)^\top s < 0$$

Thus $\bar{\sigma} > 0$ can be chosen such that

$$\frac{f(x+\sigma s) - f(x)}{\sigma} - \gamma \nabla f(x)^\top s \leq 0 \quad \forall \sigma \in (0, \bar{\sigma}].$$

■

The Armijo-inequality (7.3) is illustrated in Figure 7.1

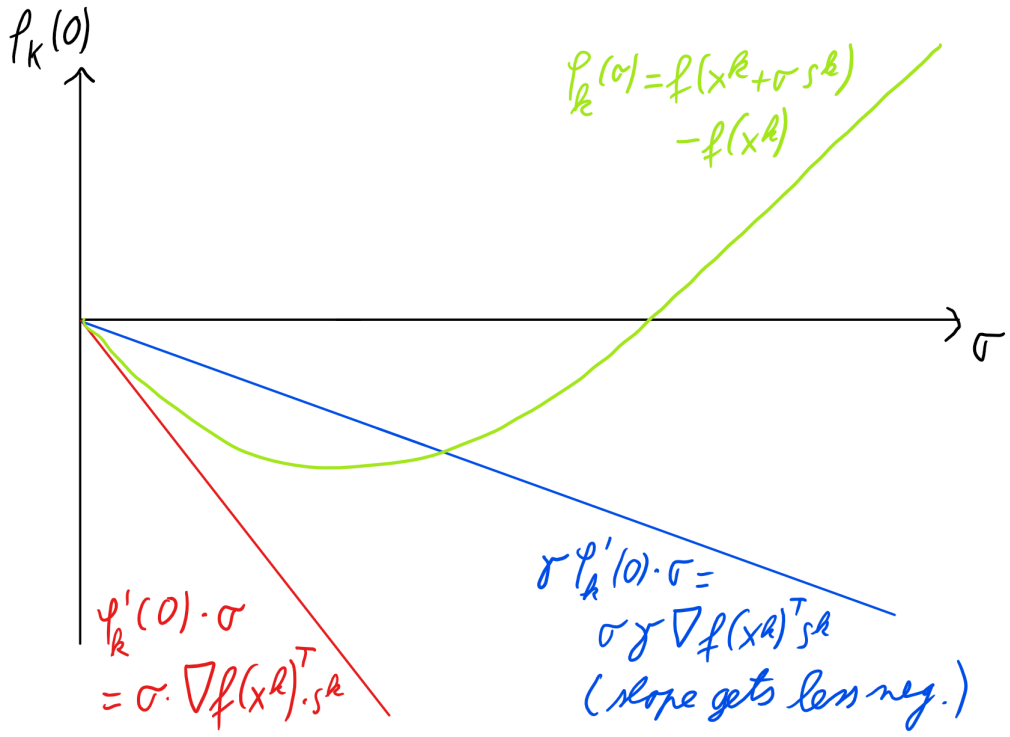


Figure 7.1.: Illustration of Armijo-inequality (7.3)

The Armijo-inequality is satisfied by all $\sigma^k = \sigma > 0$ for which the graph of $\varphi_k(\sigma)$ is on or below the line $\gamma\varphi_k'(0) \cdot \sigma$.

Algorithm 7.8 Steepest descent/gradient method

1: Choose $\beta \in (0,1)$, $\gamma \in (0,1)$, $x^0 \in \mathbb{R}^n$
2: **for** all $k = 0, 1, 2, \ldots$ **do**
3: **if** $\nabla f(x^k) = 0$ **then**
4: Stop
5: Set $s^k := -\nabla f(x^k)$
6: Compute $\sigma_k > 0$ according to the Armijo line search
7: Set $x^{k+1} = x^k + \sigma_k \cdot s^k$

In practice, the stopping criterion is $\|\nabla f(x^k)\| \leq \varepsilon$ for $\varepsilon > 0$, e. g. $\varepsilon = 10^{-6}$.

Theorem 7.9 *Let $f \in C^2(\mathbb{R}^n)$. Let $x^0 \in \mathbb{R}^n$ and let the level set $L_0 = \{x \in \mathbb{R}^n : f(x) \leq f(x^0)\}$ be compact.*
Then Algorithm 7.8 terminates with a stationary point x^k or it generates an infinite sequence $(x^k)_k$ with

a) $f(x^{k+1}) < f(x^k)$ for all k.

b) The sequence $(x^k)_k$ has at least one accumulation point.

c) Every accumulation point of $(x^k)_k$ is a stationary point.

Proof. Consider only the case, where algorithm 7.8 does not terminate after a finite number of finite steps. Theorem 7.7 states that algorithm 7.8 generates $(x^k)_k$, $(\sigma_k)_k \subseteq [0,1]$ with $\nabla f(x^k) \neq 0$ and

$$f(x^{k+1}) - f(x^k) = f(x^k + \sigma_k s^k) - f(x^k) \leq \underbrace{-\sigma_k \gamma \|\nabla f(x^k)\|^2}_{<0}$$

Hence, a) follows.
Part b)/c):
Since L_0 is compact and $f(x^k) < f(x^0)$ for all $k \geq 1$, it follows that $(x^k)_k$ has an accumulation point $\bar{x}$. Hence,

$$\lim_{l \to \infty} x^{k_l} = \bar{x} \text{ for some subsequence } (x^{k_l})_l \text{ of } (x^k)_k$$

Let $\varepsilon > 0$. Then $\|\nabla f(x^{k_l})\| \geq \varepsilon$ can only be true for a finite number of l, since otherwise f would be unbounded from below (which one can show that this is not the case).
This means that $\lim_{l \to \infty} \nabla f(x^{k_l}) = 0$.
Hence, $\nabla f(\bar{x}) \overset{\nabla f \text{ cont.}}{=} \lim_{l \to \infty} \nabla f(x^{k_l}) = 0$. ∎

Next, we show that the convergence speed of the gradient method is low even for "very nice" NLPs, i. e. strictly convex quadratic NLPs.
Let $f(x) = \frac{1}{2} x^\top A x + b^\top x + c$ with $c \in \mathbb{R}$, $b \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n \times n}$ symmetric, positive definite.

It follows that $\nabla f(x) = Ax + b$.
Hence, $\nabla f(x) = 0 \Leftrightarrow x = -A^{-1}b$.
It follows that $\nabla^2 f(x) = A$ and that f is strictly convex because of Theorem 6.4.
Consider the gradient method (algorithm 7.8) with exact line search, i. e. $\sigma_k > 0$ satisfies

$$f(x^k + \sigma_k s^k) = \min_{\sigma \geq 0} f(x^k + \sigma s^k).$$

Before we investigate the convergence speed, we illustrate the behaviour of this gradient method:
Let $L_k = \{x \in \mathbb{R}^n : f(x) = f(x^k)\}$ be the level curve to $f(x^k)$. Since $f \in C^1$ and $\nabla f(x^k) \neq 0$ it follows that L_k is a continuously differentiable hypercurve in $\mathbb{R}^n$. Let $\gamma : (-1, 1) \to L_k$ with $\gamma(0) = x^k$ be an arbitrary continuously differentiable curve in L_k. Since $f(\gamma(t)) = f(x^k)$ for all $t \in (-1, 1)$, it follows that

$$0 = (f \circ \gamma)'(0) = \nabla f(\gamma(0))^\top \cdot \gamma'(0) = \nabla f(x^k)^\top \cdot \gamma'(0)$$

Hence, s^k and $\nabla f(x^k)$ are orthogonal to the level curve (see Figure 7.2).

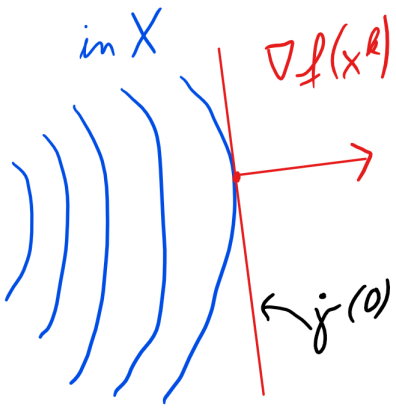


Figure 7.2.: $\nabla f(x^k)$ is orthogonal to the level curve L_k

In the exact line search, we start in x^k and follow $s^k = -\nabla f(x^k)$ until the global minimum $\sigma_k > 0$ of $\varphi(\sigma) := f(x^k + \sigma s^k)$ is reached. Hence,

$$0 = \varphi'(\sigma_k) = \nabla f(x^k + \sigma_k \cdot s^k)^\top \cdot s^k = \nabla f(x^{k+1}) \cdot s^k$$

Thus, it holds that s^k and $\nabla f(x^{k+1})$ are orthogonal. Since $\nabla f(x^{k+1})$ is the normal vector to L_{k+1}, this implies that s^k and L_{k+1} are parallel.
Thus, the algorithm moves in a "indirect" "zigzaging" way (see Figure 7.3), which is very bad if the condition

$$\kappa(\nabla^2 f(x)) = \frac{\lambda_{\max}(\nabla^2 f(x))}{\lambda_{\min}(\nabla^2 f(x))}$$

is big.

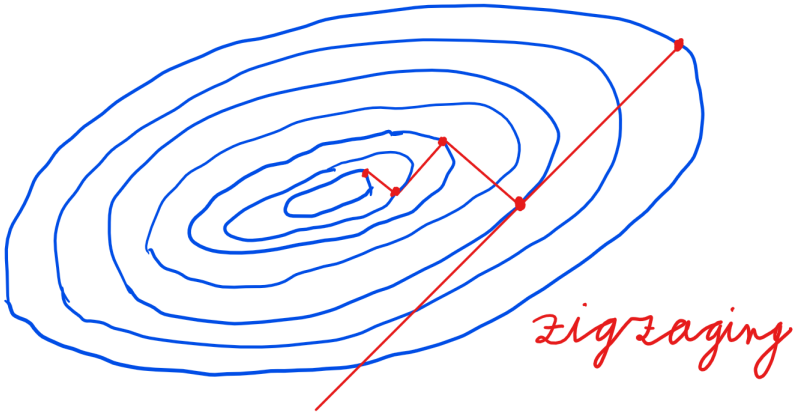


Figure 7.3.: movement of gradient method

Next, we verify this illustration formally:
The exact line search can be realized easily for this special case:
$\varphi(\sigma) := f(x^k + \sigma \cdot s^k)$
$\varphi'(\sigma) = \nabla f(x^k + \sigma s^k)^\top \cdot s^k = (b + A(x^k + \sigma s^k))^\top \cdot s^k$
$\varphi''(\sigma) = (s^k)^\top \cdot \nabla^2 f(x^k + \sigma s^k) \cdot s^k = (s^k)^\top A s^k > 0$ (since A is positive definite)
Thus, σ_k is characterized by $\varphi'(\sigma_k) = 0$.
It follows that

$$\sigma_k = -\frac{(b + Ax^k)^\top s^k}{{s^k}^\top A s^k} = -\frac{\nabla f(x^k)^\top s^k}{{s^k}^\top A s^k} = \frac{\|\nabla f(x^k)\|^2}{\nabla f(x^k)^\top A \nabla f(x^k)} = \frac{\|s^k\|^2}{{s^k}^\top A s^k}$$

Thus, σ_k can be computed explicitly once $s^k = -\nabla f(x^k)$ is known.

Theorem 7.10 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be quadratic and strictly convex. Let $(x^k)_k$ and $(\sigma_k)_k$ be generated by the gradient method with exact line search. Then it holds that*

$$f(x^{k+1}) - f(\bar{x}) \leq \left(\frac{\lambda_{\max}(A) - \lambda_{\min}(A)}{\lambda_{\max}(A) + \lambda_{\min}(A)}\right)^2 \cdot (f(x^k) - f(\bar{x})) \tag{7.4}$$

and

$$\|x^k - \bar{x}\| \leq \sqrt{\frac{\lambda_{\max}(A)}{\lambda_{\min}(A)}} \cdot \left(\frac{\lambda_{\max}(A) - \lambda_{\min}(A)}{\lambda_{\max}(A) + \lambda_{\min}(A)}\right)^k \cdot \|x^0 - \bar{x}\|, \tag{7.5}$$

where $\bar{x} := -A^{-1}b$ denotes the global minimum of f and $\lambda_{\max} = \lambda_{\max}(A)$ and $\lambda_{\min} = \lambda_{\min}(A)$ denote the maximal and minimal eigenvalues of A, respectively.

If $\kappa(A) = \frac{\lambda_{\max}}{\lambda_{\min}}$ is very large, then $\lambda_{\max} >> \lambda_{\min}$ and $\left(\frac{\lambda_{\max}(A) - \lambda_{\min}(A)}{\lambda_{\max}(A) + \lambda_{\min}(A)}\right)^2$ is close to 1, i. e. the improvement in iteration $k+1$ is little.

Lemma 7.11 (Kantorovich inequality) *Let $A \in \mathbb{R}^{n \times n}$ be symmetric and positive definite. Then it holds:*

$$\frac{\|d\|^4}{(d^\top A d)(d^\top A^{-1} d)} \geq \frac{4\lambda_{\max}(A)\lambda_{\min}(A)}{(\lambda_{\min}(A) + \lambda_{\max}(A))^2} \quad \forall d \in \mathbb{R}^n \setminus \{0\}$$

Proof of Theorem 7.10:
Since f is quadratic, Taylor in $\bar{x}$ gives

$$f(x) - f(\bar{x}) = \underbrace{\nabla f(\bar{x})(x-\bar{x})}_{=0} + \frac{1}{2}(x-\bar{x})^\top A(x-\bar{x}) = \frac{1}{2}(x-\bar{x})^\top A(x-\bar{x}) \tag{7.6}$$

and

$$\nabla f(x) = A(x-\bar{x}). \tag{7.7}$$

Taylor in x^k gives

$$f(x^{k+1}) = f(x^k) + \sigma_k \nabla f(x^k)^\top s^k + \frac{\sigma_k^2}{2}(s^k)^\top A s^k = f(x^k) - \sigma_k \|s^k\|^2 + \frac{\sigma_k^2}{2}(s^k)^\top A s^k.$$

We use $\sigma_k = \frac{\|s^k\|^2}{s^{k\top} \cdot A s^k}$ as computed before for the exact line search:

$$f(x^{k+1}) - f(\bar{x}) = f(x^k) - f(\bar{x}) - \frac{\|s^k\|^2}{s^{k\top} \cdot A s^k} \cdot \|s^k\|^2 + \frac{1}{2}\frac{\|s^k\|^4}{(s^k)^\top \cdot A s^k} = f(x^k) - f(\bar{x}) - \frac{1}{2}\frac{\|s^k\|^4}{s^{k\top} \cdot A s^k} \tag{7.8}$$

It is

$$f(x^k) - f(\bar{x}) \overset{(7.6)}{=} \frac{1}{2}(x^k - \bar{x})^\top A (x^k - \bar{x}) = \frac{1}{2}(A(x^k - \bar{x}))^\top A^{-1} (A(x^k - \bar{x})) \overset{(7.7)}{=} \frac{1}{2}(s^{k\top} \cdot A^{-1} \cdot s^k). \tag{7.9}$$

Thus,

$$\begin{aligned}
f(x^{k+1}) - f(\bar{x}) &\overset{(7.8),(7.9)}{=} \left(1 - \frac{||s^k||^4}{(s^{k\top} A s^k)(s^{k\top} A^{-1} s^k)}\right) \cdot (f(x^k) - f(\bar{x})) \\
&\overset{\text{Lemma 7.11}}{\leq} \left(1 - \frac{4\lambda_{\min}\lambda_{\max}}{(\lambda_{\max} + \lambda_{\min})^2}\right) \cdot (f(x^k) - f(\bar{x})) \\
&= \left(\frac{(\lambda_{\min} + \lambda_{\max})^2 - 4\lambda_{\min}\lambda_{\max}}{(\lambda_{\max} + \lambda_{\min})^2}\right) \cdot (f(x^k) - f(\bar{x})) \\
&= \left(\frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}}\right)^2 \cdot (f(x^k) - f(\bar{x})).
\end{aligned}$$

Moreover,

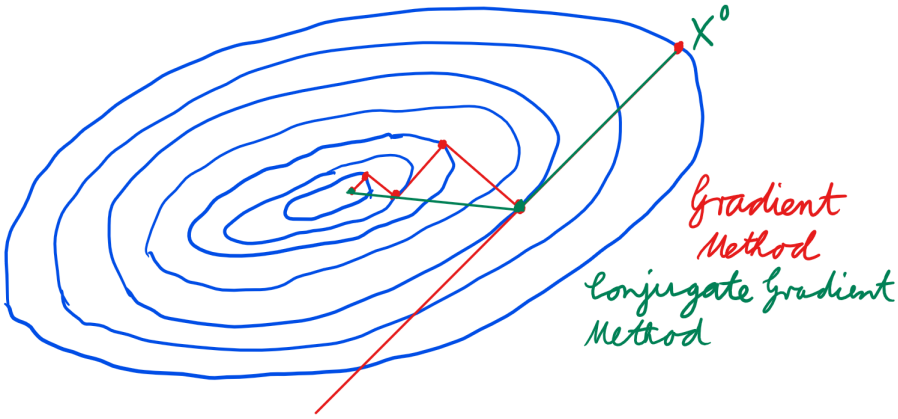
$$f(x) - f(\bar{x}) \overset{(7.6)}{=} \frac{1}{2}(x-\bar{x})^\top A (x-\bar{x}) \begin{cases} \leq \frac{\lambda_{\max}}{2}\|x-\bar{x}\|^2 \\ \geq \frac{\lambda_{\min}}{2}\|x-\bar{x}\|^2 \end{cases}. \tag{7.10}$$

Together with the already proven (7.4), it follows that

$$\begin{aligned}
\|x - \bar{x}\| &\overset{(7.10)}{\leq} \frac{2}{\lambda_{\min}}(f(x^k) - f(\bar{x})) \\
&\overset{(7.4) \text{ inductively}}{\leq} \frac{2}{\lambda_{\min}} \left(\frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}}\right)^{2k} \cdot (f(x^0) - f(\bar{x})) \\
&\overset{(7.10)}{\leq} \frac{\lambda_{\max}}{\lambda_{\min}} \left(\frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}}\right)^{2k} \|x^0 - \bar{x}\|^2
\end{aligned}$$

Taking the square root leads to (7.5).

8. Conjugate Gradient Method



Here, we consider again $f(x) = \frac{1}{2}x^\top Ax + b^\top x + c$, with $A \in \mathbb{R}^{n\times n}$ symmetric and positive definite, i. e. f is strictly convex and quadratic.
The goal of this chapter is to find a more efficient method than the gradient method for convex quadratic functions. This method can also be used for solving large systems of linear equations $Ax = -b$ and we will prove that this method terminates (in theory) after finitely many steps.

Definition 8.1 Let $A \in \mathbb{R}^{n\times n}$ be symmetric and positive definite. $s^1, \dots, s^m \in \mathbb{R}^n$ are called **A-conjugate** if

1. $s^i \neq 0 \quad \forall i = 1, \dots, m$
2. $\langle s^i, s^j \rangle_A := {s^i}^\top A s^j = 0 \quad \forall i \neq j$

Lemma 8.2

1. *If $s^0, \dots, s^m \in \mathbb{R}^n$ are **A-conjugate**, then $s^0, \dots, s^m$ are linearly independent.*
2. *If $m = n-1$, then $v = \sum_{j=0}^{n-1} \frac{{s^j}^\top Av}{{s^j}^\top As^j} s^j$ for any $v \in \mathbb{R}^n$.*

Proof. Let $v = \sum_{i=0}^{m} \alpha_i s^i$.
For $j = 0, \dots, m$, it holds:

$$ {s^j}^\top Av = {s^j}^\top A \cdot \sum_{i=0}^{m} \alpha_i s^i = \sum_{i=0}^{m} \alpha_i {s^j}^\top As^i = \alpha_j \cdot \underbrace{{s^j}^\top As^j}_{>0} $$

Hence, $\alpha_j = \frac{{s^j}^\top A v}{{s^j}^\top A s^j}$.
If $v = 0$, then $\alpha_j = 0$ for all $j = 1, \ldots, m$.
Hence, $s^0, \ldots, s^{n-1}$ are linearly independent. ∎

Theorem 8.3 *Let $f(x) = \frac{1}{2} x^\top A x + b^\top x + c$, with $A \in \mathbb{R}^{n \times n}$ symmetric and positive definite. Let $x^0 \in \mathbb{R}^n$ and $s^0, \ldots, s^{n-1}$ A-conjugate vectors.*
For $k = 0, 1, \ldots, n-1$, define $x^{k+1} := x^k + \lambda_k \cdot s^k$ with λ_k chosen by exact line search, i. e.

$$\lambda_k = \min_{\lambda \in \mathbb{R}} f(x^k + \lambda \cdot s^k).$$

Then, it holds that $f(x^n) = \min_{x \in \mathbb{R}^n} f(x)$

Proof. Let $\varphi(\lambda) = f(x^k + \lambda s^k)$.
With exact line search, it holds that

$$0 = \varphi'(\lambda_k) = {s^k}^\top \cdot \nabla f(x^{k+1}) = {s^k}^\top \cdot \left(A \left(x^0 + \sum_{j=0}^{k} \lambda_j s^j \right) + b \right) = {s^k}^\top (A x^0 + b) + \lambda_k \underbrace{{s^k}^\top A s^k}_{>0}.$$

Hence, it holds that

$$\lambda_k = \frac{-{s^k}^\top (A x^0 + b)}{{s^k}^\top A s^k}.$$ ∎

Recall that $x^n = x^0 + \sum_{k=0}^{n-1} \lambda_k s^k$. Hence, it follows that

$$\begin{aligned} x^n &= x^0 - \sum_{k=0}^{n-1} \frac{{s^k}^\top (A x^0 + b)}{{s^k}^\top A s^k} \cdot s^k \\ &= x^0 - \sum_{k=0}^{n-1} \frac{{s^k}^\top \cdot A (x^0 + A^{-1} b)}{{s^k}^\top A s^k} \cdot s^k \\ &\overset{\text{Lemma 8.2}}{=} x^0 - \Big(\underbrace{x^0 + A^{-1} b}_{\text{“}v\text{” in Lemma 8.2}} \Big) \\ &= -A^{-1} b. \end{aligned}$$

Algorithm 8.4 Conjugate Gradient Method

1: Choose starting point $x^0 \in \mathbb{R}^n$
2: Set $s^0 = -\nabla f(x^0)$
3: **for** all $k = 0, 1, 2, \ldots$ **do**
4: **if** $\nabla f(x^k) = 0$ **then**
5: Stop
6: $x^{k+1} := x^k + \lambda_k \cdot s^k$ with

$$\lambda_k = \arg\min_{\lambda \in \mathbb{R}} f(x^k + \lambda s^k)$$

7: $\gamma_{k+1} := \frac{\|\nabla f(x^{k+1})\|^2}{\|\nabla f(x^k)\|^2}$
8: $s^{k+1} := -\nabla f(x^{k+1}) + \gamma_{k+1} s^k$

Remark 8.5 The exact line search in Algorithm 8.4 is possibe. The step length λ_k satisfies

$$\begin{aligned} 0 &= \varphi'(\lambda_k) \\ &= \nabla f(x^k + \lambda_k s^k)^\top \cdot s^k \\ &= {s^k}^\top \cdot (Ax^k + b) + \lambda_k \cdot {s^k}^\top \cdot As^k \\ &= {s^k}^\top \cdot \nabla f(x^k) + \lambda_k \cdot \underbrace{{s^k}^\top \cdot As^k}_{>0 \text{ if } \nabla f(x^k) \neq 0} \end{aligned}$$

Hence, it holds that

$$\lambda_k = -\frac{{s^k}^\top \cdot \nabla f(x^k)}{{s^k}^\top A s^k}.$$

Lemma 8.6 *Let $l \in \mathbb{N}$ and $\nabla f(x^k) \neq 0$ for $k = 0, \ldots, l$, where x^k is computed as in Algorithm 8.4. Let $s^{-1} := 0$. For $0 \leq i < k \leq l$, it holds:*

a) ${s^i}^\top \cdot \nabla f(x^k) = 0$

b) ${s^k}^\top \cdot \nabla f(x^k) = -\nabla f(x^k)^\top \cdot \nabla f(x^k) = -\underbrace{\|\nabla f(x^k)\|^2}_{>0} < 0$

c) $\lambda_k > 0$

d) $\nabla f(x^i)^\top \cdot \nabla f(x^k) = 0$

e) ${s^i}^\top A s^k = 0$

Proof. The proof is done by induction on l.
$l = 0$: for a), d) and e) is nothing to show.
b) holds since $s^0 = -\nabla f(x^0)$
c) holds since

$$\lambda_0 = -\frac{{s^0}^\top \cdot \nabla f(x^0)}{{s^0}^\top \cdot As^0} = \frac{\nabla f(x^0)^\top \cdot \nabla f(x^0)}{{s^0}^\top \cdot As^0} = \frac{\overbrace{\|\nabla f(x^0)\|^2}^{>0 \text{ since } \nabla f(x^0) \neq 0}}{\underbrace{{s^0}^\top \cdot As^0}_{>0 \text{ since } A \text{ is pos. def.}}} > 0$$

$l \to l+1$: Let us assume that a)-e) hold for some $l \in \mathbb{N}$.
To show: a)-e) hold for $l+1$.
Note: if $\nabla f(x^k) \neq 0$, then by induction hypothesis c), it is $\lambda_l > 0$ and $x^{l+1} = x^l + \lambda_l \cdot s^l$ is well-defined.

a) To show: ${s^i}^\top \cdot \nabla f(x^{l+1}) = 0 \ \forall i = 0, \ldots, l$
Case 1: $i = l$: It follows that

$${s^l}^\top \cdot \nabla f(x^{l+1}) = \varphi'(\lambda_{l+1}) = 0$$

due to exact line search (where $\varphi(\lambda) = f(x^l + \lambda \cdot s^l)$).
Case 2: $i < l$:

$$\begin{aligned}
{s^i}^\top \cdot \nabla f(x^{l+1}) &= {s^i}^\top \cdot (Ax^{l+1} + b) \\
&= {s^i}^\top \cdot \left(A \left(x^{i+1} + \sum_{j=i+1}^{l} \lambda_j s^j \right) + b \right) \\
&= {s^i}^\top \cdot (Ax^{i+1} + b) + \sum_{j=i+1}^{l} \lambda_j \cdot \underbrace{{s^i}^\top \cdot As^j}_{=0 \text{ due to ind. hyp. e)}} \\
&= {s^i}^\top \cdot \nabla f(x^{i+1}) \\
&= 0 \text{ due to exact line search}
\end{aligned}$$

b)

$$\begin{aligned}
{s^{l+1}}^\top \cdot \nabla f(x^{l+1}) &\overset{\text{def.}}{=} (-\nabla f(x^{l+1}) + \gamma_{l+1} \cdot s^l)^\top \cdot \nabla f(x^{l+1}) \\
&= -\nabla f(x^{l+1}) \cdot \nabla f(x^{l+1}) + \gamma_{l+1} \cdot \underbrace{s^l \cdot \nabla f(x^{l+1})}_{=0 \text{ due to exact line search}} \\
&= -\nabla f(x^{l+1}) \cdot \nabla f(x^{l+1})
\end{aligned}$$

c) $s^{l+1} \neq 0$ due to b) and $\nabla f(x^{l+1}) \neq 0$ (stopping criterion not satisfied)
Hence, it follows from A being positive definite that ${s^{l+1}}^\top A s^{l+1} > 0$
With remark 8.5, it follows that

$$\lambda_{l+1} = -\frac{{s^{l+1}}^\top \nabla f(x^{l+1})}{{s^{l+1}}^\top \cdot As^{l+1}} \overset{b)}{=} \frac{\nabla f(x^{l+1})^\top \cdot \nabla f(x^{l+1})}{{s^{l+1}}^\top \cdot As^{l+1}} > 0$$

d) We need to show: $\nabla f(x^i)^\top \cdot \nabla f(x^{l+1}) = 0$ for $i \leq l$

$$\begin{aligned}
\nabla f(x^i)^\top \cdot \nabla f(x^{l+1}) &\overset{\text{def. } s^i}{=} (-s^i + \gamma_i s^{i-1})^\top \cdot \nabla f(x^{l+1}) \\
&= -\underbrace{{s^i}^\top \nabla f(x^{l+1})}_{=0 \text{ by a)}} + \gamma_i \underbrace{{s^{i-1}}^\top \nabla f(x^{l+1})}_{=0 \text{ by a)}} \\
&= 0
\end{aligned}$$

e) We need to show: ${s^i}^\top \cdot As^{l+1} = 0$ for $i \leq l$

$${s^i}^\top A s^{l+1} \overset{\text{def. of } s^{l+1}}{=} {s^i}^\top A(-\nabla f(x^{l+1}) + \gamma_{l+1} s^l) \tag{8.1}$$

Since $\nabla f(x^i) \neq 0$ for $i = 0, \ldots, l+1$ and since $\lambda_i > 0$ for $i = 0, \ldots, l$, it holds that

$$s^i = \frac{x^{i+1} - x^i}{\lambda_i} \tag{8.2}$$

■

Case 1: $i \leq l-1$:

$$\begin{aligned} {s^i}^\top A s^{l+1} &\overset{(8.1) \text{ and ind. hyp. e)}}{=} -{s^i}^\top \cdot A \nabla f(x^{l+1}) \\ &\overset{(8.2)}{=} -\frac{1}{\lambda_i}(Ax^{i+1} - Ax^i)^\top \cdot \nabla f(x^{l+1}) \\ &= -\frac{1}{\lambda_i}(\nabla f(x^{i+1}) - \nabla f(x^i))^\top \cdot \nabla f(x^{l+1}) \\ &\overset{d)}{=} 0 \end{aligned}$$

Case 2: $i = l$:

$$\begin{aligned} \gamma_{l+1} &= \frac{\nabla f(x^{l+1})^\top \cdot \nabla f(x^{l+1})}{\nabla f(x^l)^\top \cdot \nabla f(x^l)} \\ &\overset{d)}{=} \frac{(\nabla f(x^{l+1}) - \nabla f(x^l))^\top \cdot \nabla f(x^{l+1})}{(\nabla f(x^{l+1}) - \nabla f(x^l))^\top \cdot (-\nabla f(x^l))} \\ &\overset{a)}{=} \frac{(\nabla f(x^{l+1}) - \nabla f(x^l))^\top \cdot \nabla f(x^{l+1})}{(\nabla f(x^{l+1}) - \nabla f(x^l))^\top \cdot (-\nabla f(x^l) + \gamma_l s^{l-1})} \\ &= \frac{(A(x^{l+1} - x^l))^\top \cdot \nabla f(x^{l+1})}{(A(x^{l+1} - x^l))^\top \cdot s^l} \\ &\overset{(8.2)}{=} \frac{{s^l}^\top \cdot A \nabla f(x^{l+1})}{{s^l}^\top A s^l} \end{aligned}$$

With (8.1) we get for $i = l$:

$$\begin{aligned} {s^l}^\top \cdot A s^{l+1} &= {s^l}^\top \cdot A(-\nabla f(x^{l+1}) + \gamma_{l+1} s^l) \\ &= {s^l}^\top \cdot A \left(-\nabla f(x^{l+1}) + \frac{{s^l}^\top \cdot A \nabla f(x^{l+1})}{{s^l}^\top A s^l} \cdot s^l \right) \\ &= -{s^l}^\top \cdot A \nabla f(x^{l+1}) + {s^l}^\top A s^l \cdot \frac{{s^l}^\top A \nabla f(x^{l+1})}{{s^l}^\top A s^l} \\ &= 0 \end{aligned}$$

9. General Descent Methods

In this chapter, we consider algorithms of the following form:

Algorithm 9.1 General descent method

1: Choose a starting point $x^0 \in \mathbb{R}^n$
2: **for** all $k = 0, 1, 2, \ldots$ **do**
3: **if** $\nabla f(x^k) = 0$ **then**
4: Stop
5: Compute descent direction $s^k \in \mathbb{R}^n$
6: Compute step length $\sigma_k > 0$ with $f(x^k + \sigma_k \cdot s^k) < f(x^k)$
7: $x^{k+1} = x^k + \sigma_k \cdot s^k$

In chapter 7, we have seen that the convergence rate of the gradient method is not satisfying. In the previous chapter, we have seen that the convergence rate of the conjugate gradient method is better.
Now, we consider the following question: Under which conditions on descent directions and on step lengths can we guarantee global convergence?
With global convergence, we mean that

$$\nabla f(\bar{x}) = 0 \text{ for every accumulation point } \bar{x} \text{ of } (x^k)_k,$$

where $(x^k)_k$ is the sequence generated by algorithm 9.1.

Conditions on Descent Directions:

Definition 9.2 Let $(s^k)_k$ be the subsequence of search directions computed in algorithm 9.1. $(s^k)_k$ is called **feasible search direction** if

a) $\nabla f(x^k)^\top \cdot s^k < 0 \quad \forall k \geq 0$

b) If $\left(\frac{\nabla f(x^k)^\top \cdot s^k}{\|s^k\|}\right)_k \to 0$, then $(\nabla f(x^k))_k \to 0$

Remark 9.3

a) $\frac{\nabla f(x^k)^\top \cdot s^k}{\|s^k\|}$ is the slope of f in x in direction s^k

b) Definition 9.2 b) means: If the slope of f in x in direction s^k converges to zero, then so does the maximal slope.

c) Definition 9.2 b) can be interpreted as an abstract angle condition:

$$\frac{|\nabla f(x^k)^\top s^k|}{\|s^k\|} = \frac{-\nabla f(x^k)^\top s^k \cdot \|\nabla f(x^k)\|}{\|s^k\| \cdot \|\nabla f(x^k)\|} = \cos \sphericalangle(-\nabla f(x^k), s^k) \cdot \|\nabla f(x^k)\|$$

If the **angle condition**

$$\cos \sphericalangle(-\nabla f(x^k), s^k) = \frac{-\nabla f(x^k)^\top \cdot s^k}{\|s^k\| \cdot \|\nabla f(x^k)\|} \geq \eta$$

is fulfilled with some $0 < \eta < 1$ for all $k \geq 0$, then the condition in Definition 9.2 b) is automatically satisfied.

d) Let $\Phi : [0, \infty) \to [0, \infty)$ with $\Phi(0) = 0$ and Φ being continuous in 0.
If the **generalized angle condition**

$$\|\nabla f(x^k)\| \leq \Phi\left(\frac{-\nabla f(x^k)^\top \cdot s^k}{\|s^k\|}\right)$$

is fulfilled, then the conditon in definition 9.2 b) is also automatically satisfied.

We summarize these results in the next theorem:

Theorem 9.4 *Let $f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^1(\mathbb{R}^n)$ and $(s^k)_k$ the subsequence of descent directions generated by algorithm 9.1. Then, the following implications hold:*

1. s^k *satisfies the angle condition for all* $k \geq 0$
2. $\Rightarrow$ s^k *satisfies the generalized angle condition for all* $k \geq 0$
3. $\Rightarrow$ $(s^k)_k$ *satisfies the condition in definition 9.2 b)*

Proof.
1. $\Rightarrow$ 2.: Let $\Phi(t) := \frac{t}{\eta}$. Then, it holds that

$$\|\nabla f(x^k)\| \leq \frac{1}{\eta} \cdot \frac{-\nabla f(x^k)^\top \cdot s^k}{\|s^k\|} = \Phi\left(\frac{-\nabla f(x^k)^\top \cdot s^k}{\|s^k\|}\right)$$

2. $\Rightarrow$ 3.: Since Φ is continuous in 0 and $\Phi(0) = 0$, it holds:

$$\left(\frac{\nabla f(x^k)^\top \cdot s^k}{\|s^k\|}\right)_k \to 0 \Rightarrow \|\nabla f(x^k)\| \leq \Phi\left(\frac{-\nabla f(x^k)^\top \cdot s^k}{\|s^k\|}\right) \overset{k\to\infty}{\longrightarrow} 0 \Rightarrow (\nabla f(x^k))_k \to 0 \quad \blacksquare$$

Conditions on Step Lengths:

Definition 9.5 Let $(\sigma_k)_k$ be the subsequence of step lengths generated in algorithm 9.1. The step lengths $(\sigma_k)_k$ are called **feasible step lengths** if

a) $f(x^k + \sigma_k s^k) \leq f(x^k) \quad \forall k \geq 0$

b) $f(x^k + \sigma_k s^k) - f(x^k) \to 0 \Rightarrow \left(\frac{\nabla f(x^k)^\top \cdot s^k}{\|s^k\|}\right)_k \to 0$

Remark 9.6 We will later show that feasible step lengths ensure global convergence. For example, so-called **efficient step lengths** define feasible step lengths:

Definition 9.7 Let s^k be a descent direction of f in x^k.
$\sigma_k > 0$ is called **efficient step length** if

$$f(x^k + \sigma_k s^k) \leq f(x^k) - \theta \left(\frac{\nabla f(x^k)^\top s^k}{\|s^k\|}\right)^2$$

for some constant $\theta > 0$.

Lemma 9.8 *Let $f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^1(\mathbb{R}^n)$. Let $(x^k)_k$, $(s^k)_k$ and $(\sigma_k)_k$ be generated by algorithm 9.1. Let the condition a) in definition 9.5 be true.*
If $(\sigma_k)_k$ is a subsequence such that for all $k \geq 0$, σ_k is efficient, then $(\sigma_k)_k$ is feasible.

Proof. Let σ_k, $k \geq 0$ be an efficient step length.
Let $f(x^k + \sigma_k s^k) - f(x^k) \overset{k\to\infty}{\longrightarrow} 0$.
Then, it follows that

$$\theta \left(\frac{\nabla f(x^k)^\top \cdot s^k}{\|s^k\|}\right)^2 \leq f(x^k) - f(x^k + \sigma_k s^k) \overset{k\to\infty}{\longrightarrow} 0$$

Hence, we get that

$$\left(\frac{\nabla f(x^k)^\top s^k}{\|s^k\|}\right)_k \to 0.$$

■

Lemma 9.9 *If the Armijo line search (7.3) is applied and if, additionally,*

$$\sigma_k \geq -\alpha \frac{\nabla f(x^k)^\top s^k}{\|s^k\|^2}$$

with some constant $\alpha > 0$, then σ_k is efficient.

Proof.

$$f(x^k + \sigma_k s^k) - f(x^k) \overset{(7.3)}{\leq} \sigma_k \gamma \nabla f(x^k)^\top s^k \leq -\alpha\gamma \left(\frac{\nabla f(x^k)^\top s^k}{\|s^k\|}\right)^2$$

■

Global convergence

Theorem 9.10 *Let $f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^1(\mathbb{R}^n)$. Let algorithm 9.1 generate the sequences $(x^k)_k$, $(s^k)_k$ and $(\sigma_k)_k$. Let $\bar{x}$ be an accumulation point of $(x^k)_k$ and let $(x^{k_l})_l$ be the subsequence converging to $\bar{x}$ such that $(s^{k_l})_k$ and $(\sigma_{k_l})_l$ are feasible search directions and feasible step lengths, respectively. Then $\bar{x}$ is a stationary point.*

Proof. Let $\bar{x}$ be an accumulation point of $(x^k)_k$ with $\lim_{l\to\infty} x^{k_l} = \bar{x}$. By definition 9.5 a), it holds that $f(x^{k_l})$ is monotonically decreasing.
Hence, we get as in the proof of Theorem 7.9 that

$$\lim_{l\to\infty} f(x^{k_l}) = f(\bar{x})$$

It follows that

$$f(\bar{x}) - f(x^0) = \lim_{l\to\infty} f(x^{k_l}) - f(x^0) = \sum_{l=0}^{\infty} (f(x^{k_{l+1}}) - f(x^{k_l})).$$

Hence, it holds that

$$f(x^{k_{l+1}}) - f(x^{k_l}) \to 0$$

Since, $(\sigma_k)_k$ is feasible, it follows that

$$\left(\frac{\nabla f(x^{k_l})^\top \cdot s^{k_l}}{\|s^{k_l}\|} \right)_l \to 0 \quad \text{(see def. 9.5b))}$$

Since $(s^{k_l})_l$ is feasible we can conclude that

$$(\nabla f(x^{k_l}))_l \to 0 \quad \text{(see def. 9.2b)}$$

Due to the continuity of ∇f, it follows that

$$\nabla f(\bar{x}) = \lim_{l\to\infty} \nabla f(x^{k_l}) = 0$$

■

Armijo Rule

Next, we consider again the Armijo line search. The following example shows that the Armijo line search does not necessarilty generate feasible step lengths:

Example 9.11 Let $f : \mathbb{R} \to \mathbb{R}$, $f(x) = \frac{x^2}{8}$, $x^0 > 0$. As search directions, we choose $s^k := -2^{-k}\nabla f(x^k)$.
Once can show that the Armijo line search in algorithm 9.1 generates a monotonically descreasing sequence $(x^k)_k$ converging to some $\bar{x} \geq \frac{x^0}{2}$.
Since $f(x^k) \searrow f(\bar{x}) \geq 0$, it is $f(x^k + \sigma_k s^k) - f(x^k) \to 0$.
But $\left(\frac{\nabla f(x^k)^\top \cdot s^k}{\|s^k\|}\right)_k$ does not converge to 0.
Hence, the step lengths σ_k are not feasible (see def. 9.5b)).

Under slightly weak assumptions on the search direction, we can guarantee the feasibility of the step lengths generated by Armijo line search:

Theorem 9.12 *Let $f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^1(\mathbb{R}^n)$. Let $(x^k)_k$ be bounded. Let $\varphi : [0, \infty) \to [0, \infty)$ be a strictly monotonically increasing function such that*

$$\|s^k\| \geq \varphi\left(\frac{-\nabla f(x^k)^\top \cdot s^k}{\|s^k\|}\right) \quad \forall k \geq 0$$

holds. Then, $(\sigma_k)_k$ generated by algorithm 9.1 with Armijo line search is feasible.

Powell-Wolfe rule

On top of the Armijo rule

$$f(x^k + \sigma_k s^k) - f(x^k) \leq \gamma\sigma_k \cdot \nabla f(x^k)^\top \cdot s^k$$

the Powell-Wolf rule requires the following additional condition that ensures that $\sigma_k s^k$ is sufficiently large:

$$\nabla f(x^k + \sigma_k s^k)^\top \cdot s^k \geq \eta \cdot \nabla f(x^k)^\top \cdot s^k$$

with $0 < \gamma < \frac{1}{2}$ and $\gamma < \eta < 1$.
One can show that the Powell-Wolfe is well-defined and can be computed by some bisection search algorithm. It generates feasible step lengths under mild assumptions.

10. Newton's Method

Newton's method can be used for solving a system of nonlinear equations or for minimizing a nonlinear function. First, we consider the case of solving a system of nonlinear equations.

Newton's method for nonlinear systems of equations

Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable. Our goal is to find a solution to the system of equations

$$F(x) = 0. \tag{10.1}$$

Let us assume that $x^k \in \mathbb{R}^n$ is some iterate in the course of our algorithm for (10.1). Then, problem (10.1) is equivalent to

$$F(x^k + s) = 0, \tag{10.2}$$

where $s = d^k$ solves (10.2) if and only if $x = x^k + d^k$ solves (10.1).
The idea of Newton's method is to linearize F in x^k and use the zero of this linear approximation as the next iterate x^{k+1}.
If we develop the Taylor expansion of first order of F in x^k, we obtain that

$$F(x^k) = F(x^k) + F'(x^k) \cdot s + \rho(s) \tag{10.3}$$

$$\approx F(x^k) + F'(x^k) \cdot s \tag{10.4}$$

with $\|\rho(s)\| = o(\|s\|)$. Hence, we use the zero of (10.4) as the next approximation of (10.1). In other words, this means that we solve the system of linear equations

$$F(x^k) + F'(x^k) \cdot s = 0 \tag{10.5}$$

in order to obtain the next iterate $x^{k+1} := x^k + s^k$.
This leads to the following algorithm:

Algorithm 10.1 Newton's method for system of equations

1: Choose a starting point $x^0 \in \mathbb{R}^n$
2: **for** all $k = 0, 1, 2, \ldots$ **do**
3: **if** $F(x^k) = 0$ **then**
4: Stop
5: Compute $s^k \in \mathbb{R}^n$ by solving the Newton equation

$$F'(x^k) \cdot s^k = -F(x^k)$$

6: Set $x^{k+1} = x^k + s^k$

Next, we want to find out when and how fast Newton's method is converging to a solution of (10.1). For this purpose, we define several convergence rates:

Definition 10.2 (Convergence rates) The sequence $(x^k)_k \subseteq \mathbb{R}^n$ converges

a) **q-linearly** with rate $0 < \gamma < 1$ to $\bar{x} \in \mathbb{R}^n$, if there exists some $l \geq 0$ such that

$$\|x^{k+1} - \bar{x}\| \leq \gamma \|x^k - \bar{x}\| \quad \forall k \geq l$$

b) **q-superlinearly** to $\bar{x} \in \mathbb{R}^n$, if $x^k \to \bar{x}$ and

$$\|x^{k+1} - \bar{x}\| = o(\|x^k - \bar{x}\|) \quad \text{for } k \to \infty$$

$$\text{(Equivalently: } \frac{\|x^{k+1} - \bar{x}\|}{\|x^k - \bar{x}\|} \to 0 \quad \text{for } k \to \infty\text{)}.$$

c) **q-quadratically** to $\bar{x} \in \mathbb{R}^n$ if $x^k \to \bar{x}$ and there exists some $C > 0$ such that

$$\|x^{k+1} - \bar{x}\| \leq C \cdot \|x^k - \bar{x}\|^2 \quad \forall k \geq 0$$

d) **r-linearly** with rate $0 < \gamma < 1$ to $\bar{x} \in \mathbb{R}^n$, if there exists a sequence $(\alpha_k)_k \subseteq (0, \infty)$ which converges q-linearly with rate γ to 0 and it holds:

$$\|x^k - \bar{x}\| \leq \alpha_k \quad \text{for } k \to \infty$$

e) **r-superlinearly** to $\bar{x} \in \mathbb{R}^n$, if there exists some $(\alpha_k)_k \subseteq (0, \infty)$ which converges q-superlinearly to 0 such that

$$\|x^k - \bar{x}\| \leq \alpha_k \quad \text{for } k \to \infty$$

f) **r-quadratically** to $\bar{x} \in \mathbb{R}^n$, if there exists some $(\alpha_k)_k \subseteq (0, \infty)$ which converges q-quadratically to 0 such that

$$\|x^k - \bar{x}\| \leq \alpha_k \quad \text{for } k \to \infty$$

For proving the local convergence of Newton's method, we need the following two lemmata:

Lemma 10.3 (Banach's Lemma) *The set $\mathcal{M} \subset \mathbb{R}^{n \times n}$ of regular matrices is open and the mapping $M \in \mathcal{M} \mapsto M^{-1} \in \mathbb{R}^{n \times n}$ is a continuous mapping. Moreover, it holds for all $A \in \mathcal{M}$ and $B \in \mathbb{R}^{n \times n}$ with $\|A^{-1}B\| < 1$:*
$A + B$ is regular and it holds:

i)

$$\|(A + B)^{-1}\| \leq \frac{\|A^{-1}\|}{1 - \|A^{-1}B\|}$$

ii)

$$\|(A+B)^{-1} - A^{-1}\| \leq \frac{\|A^{-1}\| \cdot \|A^{-1}B\|}{1 - \|A^{-1}B\|}$$

Proof. Let $A \in \mathcal{M}$, $B \in \mathbb{R}^{n\times n}$ with $\|A^{-1}B\| < 1$. Let $M := -A^{-1}B$ and $M^0 := I$, where $I \in \mathbb{R}^{n\times n}$ is the identity matrix. Then it holds that

$$S = \sum_{k=0}^{\infty} M^k$$

converges since it holds for $S_n := \sum_{k=0}^{n} M^k$ that

$$\|S - S_n\| = \|\sum_{k=n+1}^{\infty} M^k\| \leq \sum_{k=n+1}^{\infty} \|M\|^k = \|M\|^{n+1} \cdot \sum_{k=0}^{\infty} \|M\|^k = \frac{\|M\|^{n+1}}{1 - \|M\|} \overset{n\to\infty}{\longrightarrow} 0$$

Moreover, it holds that

$$S_n(I - M) = (I - M)S_n = (I - M)\sum_{k=0}^{n} M^k = I - M^{n+1}$$

Taking the limit $n \to \infty$ leads to $S(I - M) = (I - M)S = I$.
Hence, it follows that $(I - M) \in \mathcal{M}$ and $(I - M)^{-1} = S$.
Since $A + B = A(I + A^{-1}B) = A(I - M)$, it holds that $(A + B) \in \mathcal{M}$ and

$$(A+B)^{-1} = SA^{-1} \tag{10.6}$$

It follows that

$$\|(A+B)^{-1}\| \overset{(10.6)}{\leq} \|A^{-1}\| \cdot \|S\| \leq \|A^{-1}\| \cdot \sum_{k=0}^{\infty} \|M\|^k = \frac{\|A^{-1}\|}{1 - \|M\|}$$

and that

$$\begin{aligned}
\|(A+B)^{-1} - A^{-1}\| &\overset{(10.6)}{=} \|SA^{-1} - A^{-1}\| && (10.7)\\
&= \|\sum_{k=0}^{\infty} M^k A^{-1} - A^{-1}\| && (10.8)\\
&\leq \|A^{-1}\| \cdot \sum_{k=1}^{\infty} \|M\|^k && (10.9)\\
&\leq \frac{\|A^{-1}\| \cdot \|M\|}{1 - \|M\|} && (10.10)
\end{aligned}$$

If $\|B\| \to 0$ then $\|M\| \to 0$. Hence, it follows with (10.10) and the epsilon-delta criterion that taking the inverse of a matrix is a continuous mapping. ∎

In the next lemma, we show that a solution $\bar{x}$ of (10.1) is isolated (i. e. there is no other solution in a neighborhood around $\bar{x}$), if $F'(\bar{x})$ is regular:

Lemma 10.4 *Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable. Let $F(\bar{x}) = 0$ and $F'(\bar{x})$ be regular. Then there exists some $\varepsilon > 0$ and $\gamma > 0$ with*

$$\|F(x)\| \geq \gamma \|x - \bar{x}\| \quad \forall x \in B_\varepsilon(\bar{x}).$$

In particular, this means that $\bar{x}$ is an isolated zero of F.

Proof. It holds that

$$\|x - \bar{x}\| = \|F'(\bar{x})^{-1} F'(\bar{x})(x - \bar{x})\| \leq \|F'(\bar{x})^{-1}\| \cdot \|F'(\bar{x})(x - \bar{x})\|.$$

Let $\gamma = \frac{1}{2\|F'(\bar{x})\|}$. Then we get that

$$\|F'(\bar{x})(x - \bar{x})\| \geq 2\gamma \|x - \bar{x}\|. \tag{10.11}$$

By the definition of differentiability, there exists some $\varepsilon > 0$ such that

$$\|F(x) - F(\bar{x}) - F'(x)(x - \bar{x})\| \leq \gamma \|x - \bar{x}\| \quad \forall x \in B_\varepsilon(\bar{x}). \tag{10.12}$$

■

It follows that

$$\begin{aligned} 2\gamma\|x - \bar{x}\| &\overset{(10.11)}{\leq} \|F'(\bar{x})(x - \bar{x})\| \\ &= \|F(x) - (F(x) - F'(\bar{x})(x - \bar{x}))\| \\ &\overset{F(\bar{x})=0}{\leq} \|F(x)\| + \|F(x) - F(\bar{x}) - F'(\bar{x})(x - \bar{x})\| \\ &\leq \|F(x)\| + \gamma\|x - \bar{x}\| \end{aligned}$$

Next, we can prove the local quadratic convergence of Newton's method:

Theorem 10.5 (Local convergence of Newton's method) *Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable, $\bar{x} \in \mathbb{R}^n$ with $F(\bar{x}) = 0$ and F' invertible in $\bar{x}$.*
Then there exists some $\delta > 0$ and $C > 0$ such that it holds:

a) $\bar{x}$ is the only zero of F on $B_\delta(\bar{x})$.

b) $\|F'(x)^{-1}\| \leq C$ for all $x \in B_\delta(\bar{x})$

c) Algorithm 10.1 terminates for all $x^0 \in B_\delta(\bar{x})$ either with $x^k = \bar{x}$ or it generates a sequence $(x^k)_k \subset B_\delta(\bar{x})$ which converges q-superlinearly to $\bar{x}$.

d) If F' is Lipschitz continuous on $B_\delta(\bar{x})$ with constant L, i. e.

$$\|F'(x) - F'(y)\| \leq L \cdot \|x - y\| \quad \forall x, y \in B_\delta(\bar{x}),$$

then the rate of convergence is even q-quadratic:

$$\|x^{k+1} - \bar{x}\| \leq \frac{C \cdot L}{2} \cdot \|x^k - \bar{x}\|^2 \quad \forall k \geq 0.$$

Proof.

a) Lemma 10.4 says that there exists some $\delta_1 > 0$, such that $\bar{x}$ is the only zero of F on $B_{\delta_1}(\bar{x})$.

b) Since F' is continuous and $F'(\bar{x})$ is regular, there exist $0 < \delta_2 \leq \delta_1$ and $C > 0$ with $\|F'(x)^{-1}\| \leq C$ for all $x \in B_{\delta_2}(\bar{x})$.

c) It holds that

$$F(y) = F(x) + F'(x)(y-x) + R(x,y) \quad \forall x, y \in \mathbb{R}^n \tag{10.13}$$

with

$$R(x,y) = \int_0^1 F'(x + t(y-x))(y-x)dt - F'(x)(y-x). \tag{10.14}$$

This follows from analysis, because it holds with $G(t) := F(x + t(y-x))$ that $G'(t) = F'(x + t(y-x)) \cdot (y-x)$ and

$$\int_0^1 F'(x + t(y-x))(y-x)dt = \int_0^1 G'(t)dt = G(1) - G(0) = F(y) - F(x).$$

For $x^k \in B_{\delta_2}(\bar{x})$, it follows that

$$\begin{aligned} x^{k+1} - \bar{x} &= x^{k+1} - x^k + x^k - \bar{x} \overset{\text{Newton step}}{=} -F'(x^k)^{-1}F(x^k) + x^k - \bar{x} \\ &= F'(x^k)^{-1}(-F(x^k) + F'(x^k)(x^k - \bar{x})) \\ &\overset{F(\bar{x})=0}{=} F'(x^k)^{-1}(F(\bar{x}) - F(x^k) - F'(x^k)(\bar{x} - x^k)) \\ &\overset{(10.13)}{=} F'(x^k)^{-1}R(x^k, \bar{x}). \end{aligned} \tag{10.15}$$

We can conclude that

$$\begin{aligned} \|R(x,\bar{x})\| &\overset{(10.14)}{\leq} \int_0^1 \|(F'(x + t(\bar{x} - x)) - F'(x))(\bar{x} - x)\|dt \\ &\leq \int_0^1 \|F'(x + t(\bar{x} - x)) - F'(x)\|dt \|\bar{x} - x\|. \end{aligned} \tag{10.16}$$

It follows from the continuity of F' that

$$\int_0^1 \|F'(x + t(\bar{x} - x)) - F'(x)\|dt \overset{x \to \bar{x}}{\longrightarrow} 0. \tag{10.17}$$

Hence, it follows with (10.16) that for any $0 < \alpha < 1$, there exists some $0 < \delta \leq \delta_2$ such that

$$\|R(x,\bar{x})\| \leq \frac{\alpha}{C}\|x - \bar{x}\| \quad \forall x \in B_\delta(\bar{x}).$$

For any $x^k \in B_\delta(\bar{x})$, this leads to

$$\begin{aligned} \|x^{k+1} - \bar{x}\| &\overset{(10.15)}{=} \|F'(x^k)^{-1}R(x^k,\bar{x})\| \leq \|F'(x^k)^{-1}\| \cdot \|R(x^k,\bar{x})\| \\ &\overset{\text{b)}}{\leq} C\frac{\alpha}{C}\|x^k - \bar{x}\| = \alpha\|x^k - \bar{x}\|. \end{aligned} \tag{10.18}$$

This shows that for $x^0 \in B_\delta(\bar{x})$, it holds that $x^1 \in B_{\alpha\delta}(\bar{x}) \subset B_\delta(\bar{x})$ and, inductively, that $x^k \in B_{\alpha^k\delta}(\bar{x}) \subset B_\delta(\bar{x})$. If $F(x^k) = 0$, then algorithm 10.1 terminates. Because of $x^k \in B_\delta(\bar{x}) \subset B_{\delta_1}(\bar{x})$ and a), it must be $x^k = \bar{x}$. Hence, algorithm 10.1 terminates with $x^k = \bar{x}$ or it generates a sequence $(x^k)_k$ converging to $\bar{x}$.
The q-superlinear convergence follows from

$$\frac{\|x^{k+1} - \bar{x}\|}{\|x^k - \bar{x}\|} \overset{(10.15),(10.16)}{\leq} \|F(x^k)^{-1}\| \int_0^1 \|F'(x^k + t(\bar{x} - x^k)) - F'(x^k)\| dt$$
$$\overset{b),(10.17)}{\leq} C \int_0^1 \|F'(x^k + t(\bar{x} - x^k)) - F'(x^k)\| dt \overset{k\to\infty}{\longrightarrow} 0.$$

d) Since F' is Lipschitz-continuous on $B_\delta(\bar{x})$, it follows that

$$\int_0^1 \|F'(x^k + t(\bar{x} - x^k)) - F'(x^k)\| dt \leq \int_0^1 L \cdot t \|x^k - \bar{x}\| dt = \frac{L}{2} \|x^k - \bar{x}\| \quad (10.19)$$

Hence, it follows that

$$\|x^{k+1} - \bar{x}\| \overset{(10.15)}{\leq} \|F'(x^k)^{-1}\| \cdot \|R(x^k, \bar{x})\| \overset{b),(10.16),(10.19)}{\leq} \frac{C \cdot L}{2} \|x^k - \bar{x}\|^2 \qquad \blacksquare$$

Next, we want to solve an optimization problem by using Newton's method.

Newton's method for optimization problems

Let $f : \mathbb{R}^n \to \mathbb{R}$ with $f \in C^2(\mathbb{R}^n)$, i. e. $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$ is continuously differentiable. The optimization problem is, as usual

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

A first approach is to use the first-order necessary condition for local minima. This means that we apply Newton's method (algorithm 10.1) to $\nabla f(x) = 0$. This leads to the following algorithm:

Algorithm 10.6 Newton's method for optimization problems

1: Choose a starting point $x^0 \in \mathbb{R}^n$
2: **for** all $k = 0, 1, 2, \ldots$ **do**
3: **if** $\nabla f(x^k) = 0$ **then**
4: Stop
5: Compute $s^k \in \mathbb{R}^n$ by solving the Newton equation

$$\nabla^2 f(x^k) \cdot s^k = -\nabla f(x^k)$$

6: Set $x^{k+1} = x^k + s^k$

A second approach is to approximate f by a quadratic function, the taylor polynomial of second order in the current point x^k:

$$f(x^k + s) \approx \underbrace{f(x^k) + \nabla f(x^k)^\top s + \frac{1}{2} s^\top \nabla^2 f(x^k) s}_{=:q_k(s)}$$

The next iteration point x^{k+1} is then chosen as the stationary point of $\nabla q_k(s)$. By setting $\nabla q_k(s) = 0$, we obtain the equation

$$\nabla f(x^k) + \nabla^2 f(x^k) \cdot s = 0$$

and the next direction $s^k = -\nabla^2 f(x^k)^{-1} \cdot \nabla f(x^k)$.
Hence, we obtain again algorithm 10.6. Similarly to Theorem 10.5, we obtain the following convergence result:

Theorem 10.7 (Local convergence of Newton's method) *Let $f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^2(\mathbb{R}^n)$ and $\bar{x} \in \mathbb{R}^n$ a local minimum that satisfies the second-order sufficient conditions. Then there exist some $\delta > 0$ and $\mu > 0$ such that the following holds:*

a) $\bar{x}$ is the only stationary point

b) $\lambda_{\min}(\nabla^2 f(x)) \geq \mu$ for all $x \in B_\delta(\bar{x})$

c) Algorithm 10.6 terminates with $x^k = \bar{x}$ or it generates a sequence $(x^k)_k \subset B_\delta(\bar{x})$ that converges q-superlinearly to $\bar{x}$.

d) If $\nabla^2 f$ is Lipschitz-continuous on $B_\delta(\bar{x})$, then the convergence rate is even q-quadratic.

Proof. The proof is similar to the one of Theorem 10.5. ∎

Example 10.8 Let $f : \mathbb{R} \to \mathbb{R}$, $f(x) = \sqrt{x^2 + 1}$. Then, it follows that

$$\nabla f(x) = \frac{x}{\sqrt{x^2+1}} \text{ and } \nabla^2 f(x) = \frac{1}{\sqrt{x^2+1}} - \frac{x^2}{(x^2+1)^{\frac{3}{2}}} = \frac{1}{(x^2+1)^{\frac{3}{2}}}.$$

The Newton equation is given by

$$\frac{1}{((x^k)^2+1)^{\frac{3}{2}}} \cdot s^k = -\frac{x^k}{\sqrt{(x^k)^2+1}}$$

and it follows that $s^k = -x^k((x^k)^2 + 1)$ and $x^{k+1} = x^k + s^k = -(x^k)^3$. This implies

- convergence for $|x^0| < 1$
- divergence for $|x^0| > 1$
- alternation for $|x^0| = 1$.

This means that Newton's method does not converge for $|x^0| \geq 1$.

Globalized Newton's Method

The last example showed that Newton's method does, in general, not converge globally. Hence, the next idea is to derive a "globalized Newton's method" by making use of chapter 9, Theorem 9.10, feasible step lengths and feasible search directions. Note that Theorem 9.10 is applicable if the search directions and step lengths are feasible. This can be ensured by using a **generalized angle condition**. If this condition is satisfied by the Newton step, then we do a Newton step. Otherwise, we a do a "gradient step". One possible generalized angle condition is used in the following globalized Newton's method:

Algorithm 10.9 A globalized Newton's method

1: Choose $x^0 \in \mathbb{R}^n$, $\beta \in (0,1)$, $\gamma \in (0,1)$, $\alpha_1, \alpha_2 > 0$, $p > 0$
2: **for** all $k = 0, 1, 2, \ldots$ **do**
3: **if** $\nabla f(x^k) = 0$ **then**
4: Stop
5: Compute $d^k \in \mathbb{R}^n$ by solving the Newton equation

$$\nabla^2 f(x^k) \cdot d^k = -\nabla f(x^k)$$

6: **if** d_k satisfies

$$-\nabla f(x^k)^\top \cdot d^k \geq \min\{\alpha_1, \alpha_2 \cdot \|d^k\|^p\} \cdot \|d^k\|^2$$

then
7: set $s^k = d^k$
8: **else**
9: set $s^k = -\nabla f(x^k)$.
10: Compute $\sigma_k > 0$ by Armijo line search (7.3)
11: Set $x^{k+1} = x^k + s^k$

The next theorem states the global convergence of this algorithm:

Theorem 10.10 *Let $f : \mathbb{R}^n \to \mathbb{R}$ with $f \in C^2(\mathbb{R}^n)$. Then Algorithm 10.9 terminates with $\nabla f(x^k) = 0$ or it generates a sequence $(x^k)_k$ such that every accumulation point of $(x^k)_k$ is a stationary point of f.*

Proof. The idea of the proof is to shows the following three statements and use Theorem 9.10:

1. $(s^k)_k$ are descent directions.
2. $(s^k)_k$ are feasible search directions.
3. (σ_k) are feasible step length.

Here, we only show 1.:
Let K_g be the set of all $k \geq 0$ such that $s^k = -\nabla f(x^k)$.

Let K_n be the set of all $k \geq 0$ such that $s^k = d^k \neq -\nabla f(x^k)$.
It follows that

$$\frac{-\nabla f(x^k)^\top \cdot s^k}{\|s^k\|} = \|\nabla f(x^k)\| \quad \forall k \in K_g.$$

For all $k \in K_n$, it holds that $s^k = d^k = -\nabla^2 f(x^k)^{-1} \cdot \nabla f(x^k) \neq 0$ and that

$$\frac{-\nabla f(x^k)^\top \cdot s^k}{\|s^k\|} \geq \min\{\alpha_1, \alpha_2 \cdot \|s^k\|^p\} \cdot \|s^k\| > 0.$$

Hence, s^k is a descent direction for all $k \geq 0$. ∎

11. Newton-like Methods

In Newton's method, we need to solve the Newton equation

$$F'(x^k)s^k = -F(x^k)$$

or (in case of an optimization problem)

$$\nabla^2 f(x^k)s^k = -\nabla f(x^k)$$

in every iteration. In practice, computing $F(x)$ (or $\nabla f(x)$) can be very demanding. Computing the Jacobian matrix $F'(x)$ (or the Hessian matrix $\nabla^2 f(x)$) can be even more demanding or sometimes even impossible. Additionally, the Newton-step is not always well-defined and, for the case of an optimization problem, not always a descent direction. In practice, $F'(x^k)$ (or $\nabla^2 f(x^k)$) is therefore often approximated by some matrix $M_k \in \mathbb{R}^{n\times n}$. Such algorithms are called **Newton-like methods**.
The goal in this chapter is to study conditions under which Newton-like methods converge with q-superlinear speed. Newton-like methods have the following general structure:

Algorithm 11.1 General outline of Newton-like methods

1: Choose $x^0 \in \mathbb{R}^n$
2: **for** all $k = 0, 1, 2, \ldots$ **do**
3: **if** $F(x^k) = 0$ **then**
4: Stop
5: Compute $s^k \in \mathbb{R}^n$ by solving $M_k \cdot s^k = -F(x^k)$
6: Set $x^{k+1} = x^k + s^k$

For studying q-superlinear convergence, we need the following lemma:

Lemma 11.2 *Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable. Let $X \subseteq \mathbb{R}^n$ be compact and convex. Then F is Lipschitz-continuous on X with Lipschitz-constant*

$$L = \max_{x\in X} \|F'(x)\|.$$

Proof. Since X is compact and $\|F'\|$ is continuous, it follows that $\|F'\|$ attains its maximum $L = \max_{x\in X}\|F'(x)\|$ on X. ∎

In a similar way as in the proof of Theorem 10.5, it follows that

$$\begin{aligned}\|F(y) - F(x)\| &= \|\int_0^1 F'(x + t(y-x)) \cdot (y-x)dt\| \\ &\leq \int_0^1 \|F'(x + t(y-x)) \cdot (y-x)\|dt \\ &\leq \int_0^1 \|F'(x + t(y-x))\|dt \cdot \|y-x\| \\ &\leq L \cdot \|y-x\|.\end{aligned}$$

The next theorem states equivalences to q-superlinear convergence:

Theorem 11.3 *Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable and let $\bar{x} \in \mathbb{R}^n$ such that $F'(\bar{x})$ is regular. Let $(x^k)_k$ be a sequence converging to $\bar{x}$ with $x^k \neq \bar{x}$ for all $k \in \mathbb{N}$. Then, the following statements are equivalent:*

a) $(x^k)_k$ *converges q-superlinearly to* $\bar{x}$ *and* $F(\bar{x}) = 0$.

b) $\|F(x^k) + F'(\bar{x})(x^{k+1} - x^k)\| = o(\|x^{k+1} - x^k\|)$.

c) $\|F(x^k) + F'(x^k)(x^{k+1} - x^k)\| = o(\|x^{k+1} - x^k\|)$.

Proof. It holds that

$$\begin{aligned}F(x^{k+1}) &= F(x^k) + \int_0^1 F'(x^k + t(x^{k+1} - x^k))(x^{k+1} - x^k)dt \\ &= \int_0^1 (F'(x^k + t(x^{k+1} - x^k)) - F'(\bar{x}))(x^{k+1} - x^k)dt \\ &\quad + F(x^k) + F'(\bar{x})(x^{k+1} - x^k).\end{aligned} \tag{11.1}$$

"*b*) ⇒ *a*):" It follows from (11.1) and b) that

$$\begin{aligned}\|F(x^{k+1})\| &\leq \int_0^1 \|F'(x^k + t(x^{k+1} - x^k)) - F'(\bar{x})\|dt\|x^{k+1} - x^k\| \\ &\quad + \underbrace{\|F(x^k) + F'(\bar{x})(x^{k+1} - x^k)\|}_{=o(\|x^{k+1}-x^k\|) \text{ by Assumption}} \\ &= o(\|x^{k+1} - x^k\|),\end{aligned}$$

since $x^k \to \bar{x}$.
Hence, there exists a sequence $(\varepsilon_k)_k \subseteq (0, \infty)$ converging to 0 and some $l \geq 0$ with

$$\|F(x^{k+1})\| \leq \varepsilon_k \cdot \|x^{k+1} - x^k\| \quad \forall k \geq l. \tag{11.2}$$

From $x^{k+1} \to \bar{x}$, it follows that $F(\bar{x}) = \lim_{k\to\infty} F(x^{k+1}) = 0$.
Since $F'(\bar{x})^{-1}$ exists, it follows by Lemma 10.4 that there exists some $\gamma > 0$ such that

$$\|F(x^{k+1})\| \geq \gamma \cdot \|x^{k+1} - \bar{x}\| \quad \text{for large } k. \tag{11.3}$$

For sufficiently large k, it is $\varepsilon_k \leq \frac{\gamma}{2}$ and

$$\begin{aligned}
\|x^{k+1} - \bar{x}\| &\overset{(11.3)}{\leq} \frac{1}{\gamma}\|F(x^{k+1})\| \\
&\overset{(11.2)}{\leq} \frac{\varepsilon_k}{\gamma}\|x^{k+1} - x^k\| \\
&= \frac{\varepsilon_k}{\gamma}\|x^{k+1} - \bar{x} + \bar{x} - x^k\| \\
&\leq \frac{\varepsilon_k}{\gamma}\|x^{k+1} - \bar{x}\| + \frac{\varepsilon_k}{\gamma}\|x^k - \bar{x}\| \\
&\leq \frac{1}{2}\|x^{k+1} - \bar{x}\| + \frac{\varepsilon_k}{\gamma}\|x^k - \bar{x}\|.
\end{aligned}$$

Hence, it follows that

$$\|x^{k+1} - \bar{x}\| \leq \frac{2\varepsilon_k}{\gamma}\|x^k - \bar{x}\| = o(\|x^k - \bar{x}\|).$$

"$a) \Rightarrow b)$:" Because of a), there exists some $l \geq 0$ such that

$$\|x^k - \bar{x}\| \leq \|x^{k+1} - x^k\| + \|x^{k+1} - \bar{x}\| \overset{a)}{\leq} \|x^{k+1} - x^k\| + \frac{1}{2}\|x^k - \bar{x}\| \quad \forall k \geq l.$$

Hence, it holds that

$$\|x^k - \bar{x}\| \leq 2\|x^{k+1} - x^k\| \quad \forall k \geq l. \tag{11.4}$$

Since $(x^k)_k$ converges to $\bar{x}$, it holds that $(x^k)_k$ lies in a sufficiently large compact ball $\overline{B_\varepsilon(\bar{x})}$. With lemma 11.2, it follows that there exists some $L > 0$ with

$$\|F(x^{k+1})\| = \|F(x^{k+1}) - F(\bar{x})\| \leq L\|x^{k+1} - \bar{x}\| \quad \forall k \geq 0. \tag{11.5}$$

It follows that

$$\begin{aligned}
&\|F(x^k) + F'(\bar{x})(x^{k+1} - x^k)\| \\
&\overset{(11.1)}{\leq} \|F(x^{k+1})\| + \int_0^1 \|F'(x^k + t(x^{k+1} - x^k)) - F'(\bar{x})\|dt\|x^{k+1} - x^k\| \\
&\overset{(11.5)}{=} \mathcal{O}(\|x^{k+1} - \bar{x}\|) + o(\|x^{k+1} - x^k\|) \\
&\overset{a)}{=} o(\|x^k - \bar{x}\|) + o(\|x^{k+1} - x^k\|) \\
&\overset{(11.4)}{=} o(\|x^{k+1} - x^k\|).
\end{aligned}$$

"$b) \Rightarrow c)$:" Because of b) and $x^k \to \bar{x}$, it holds that

$$\begin{aligned}
&\|F(x^k) + F'(x^k)(x^{k+1} - x^k)\| \\
&\leq \|F(x^k) + F'(\bar{x})(x^{k+1} - x^k)\| + \|F'(x^k) - F'(\bar{x})\| \cdot \|x^{k+1} - x^k\| \\
&= o(\|x^{k+1} - x^k\|) + o(\|x^{k+1} - x^k\|) \\
&= o(\|x^{k+1} - x^k\|).
\end{aligned}$$

"$c) \Rightarrow b)$:" Because of c) and $x^k \to \bar{x}$, it follows that

$$\begin{aligned}
&\|F(x^k) + F'(\bar{x})(x^{k+1} - x^k)\| \\
&\leq \|F(x^k) + F'(x^k)(x^{k+1} - x^k)\| + \|F'(\bar{x}) - F'(x^k)\| \cdot \|x^{k+1} - x^k\| \\
&= o(\|x^{k+1} - x^k\|) + o(\|x^{k+1} - x^k\|) \\
&= o(\|x^{k+1} - x^k\|).
\end{aligned}$$

■

Next, we consider the special sequence $(x^k)_k$ that is generated by algorithm 11.1:

Corollary 11.4 (Dennis-Moré condition) *Let $(x^k)_k$ be generated by algorithm 11.1. Let $x^k \to \bar{x}$ and let $F'(\bar{x})$ be regular.*
Then, the following statements are equivalent:

a) $(x^k)_k$ *converges q-superlinearly to* $\bar{x}$ *and* $F(\bar{x}) = 0$.

b) $\|(M_k - F'(\bar{x}))(x^{k+1} - x^k)\| = o(\|x^{k+1} - x^k\|)$

c) $\|(M_k - F'(x^k))(x^{k+1} - x^k)\| = o(\|x^{k+1} - x^k\|)$

Proof. Note that $M_k s^k = -F(x^k)$ and $x^{k+1} - x^k = s^k$. Hence, it holds that

$$\|(M_k - F'(\bar{x}))(x^{k+1} - x^k)\| = \|F(x^k) + F'(\bar{x})(x^{k+1} - x^k)\|,$$

i. e. Corollary 11.4b) is equivalent to Theorem 11.3b) and Corollary 11.4c) is equvialent to Theorem 11.3. Hence, the statement follows from Theorem 11.3. ■

The Dennis-Moré condition states that for q-superlinear convergence, it is sufficient that $M_k s^k$ and $F'(\bar{x})s^k$ (or: $F'(x^k)s^k$) are sufficiently close, i. e.

$$M_k s^k \approx F'(\bar{x})s^k,$$

which is equivalent to $\|M_k s^k - F'(\bar{x})s^k\| = o(\|s^k\|)$.
If $M_k \to F'(\bar{x})$, then

$$\|(M_k - F'(\bar{x}))(x^{k+1} - x^k)\| \leq \|M_k - F'(\bar{x})\| \cdot \|x^{k+1} - x^k\| = o(\|x^{k+1} - x^k\|).$$

Thus, superlinear convergence follows in this case.

Next, we consider a "globalized" Newton-like method (compare chapter 10) in the case of an optimization problem:

Algorithm 11.5 A globalized Newton-like method

Let $x^0 \in \mathbb{R}^n$, $\beta \in (0,1)$, $\gamma \in (0,1)$, $\alpha_1, \alpha_2 > 0$, $p > 0$
for $k = 0, 1, 2, \ldots$ **do**
  if $\nabla f(x^k) = 0$ **then**
    Stop
  Choose regular matrix $M_k \in \mathbb{R}^{n \times n}$
  Compute $d^k \in \mathbb{R}^n$ by solving $M_k \cdot d^k = -\nabla f(x^k)$
  if d^k satisfies $-\nabla f(x^k)^\top d^k \geq \min\{\alpha_1, \alpha_2 \cdot \|d^k\|^p\} \cdot \|d^k\|^2$ **then**
    $s^k := d^k$
  else
    $s^k := -\nabla f(x^k)$
  Compute $\sigma_k > 0$ by Armijo line search
  Set $x^{k+1} = x^k + \sigma_k s^k$

The proof of Theorem 10.10 works here as well (if $(\|M_k\|)_k$ is bounded), i. e. global convergence is guaranteed for Algorithm 11.5.

12. Inexact Newton Methods

The idea in this chapter is to solve the Newton equation

$$F'(x^k)s^k = -F(x^k)$$

approximately. In the original Newton's method, a system $M \cdot s = b$ with $M = F'(x^k)$ and $b = -F(x^k)$ is solved. For optimization problems, it holds that $M = \nabla^2 f(x^k)$, $b = -\nabla f(x^k)$ and that M is symmetric. Additionally, M is even positive definite, if x^k is located close enough to a local minimum that fulfills the second-order sufficient optimality conditions. In practice, an iterative solver, as e. g. the conjugate gradient method, is used to solve the Newton equation and the iterative solver stops if $\|Ms^k - b\|$ is sufficiently small. **Inexact Newton methods** have the following general structure:

Algorithm 12.1 General outline of inexact Newton method

1: Let $x^0 \in \mathbb{R}^n$
2: **for** $k = 0, 1, 2, \ldots$ **do**
3: **if** $F(x^k) = 0$ **then**
4: Stop
5: Compute $s^k \in \mathbb{R}^n$ by solving $F'(x^k)s^k = -F(x^k)$ approximately
6: Set $x^{k+1} = x^k + s^k$

To ensure that the Newton equation is solved "sufficiently good" in the inexact Newton methods, we assume that s^k satisfies

$$\|F(x^k) + F'(x^k)s^k\| \leq \eta_k \|F(x^k)\|, \tag{12.1}$$

where $\eta_k > 0$ is sufficiently small.
The next theorem states the convergence rate of inexact Newton methods:

Theorem 12.2 *Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable. Let $\bar{x} \in \mathbb{R}^n$ with $F(\bar{x}) = 0$ and let $F'(\bar{x})$ be regular.*
Then there exist $\varepsilon > 0$ and $\eta > 0$ such that the following holds:

a) If $x^0 \in B_\varepsilon(\bar{x})$ and assumption (12.1) is fulfilled with $\eta_k \leq \eta$ for all k, then algorithm 12.1 terminates with $x^k = \bar{x}$ or it generates a sequence $(x^k)_k$ which converges q-linearly to $\bar{x}$.

b) If $\eta_k \to 0$, then the convergence rate is q-superlinear.

c) *If $\eta_k \in \mathcal{O}(\|F(x^k)\|)$ and if F' is Lipschitz-continuous on $B_\varepsilon(\bar{x})$, then the convergence rate is q-quadratic.*

Proof. We only prove b):
The idea of the proof is to apply Theorem 11.3. By assumption (12.1), it holds that

$$\|F(x^k) + F'(x^k)(x^{k+1} - x^k)\| = \|F(x^k) + F'(x^k)s^k\| \leq \eta_k \|F(x^k)\| \tag{12.2}$$

It follows from a) that $x^k \to \bar{x}$ and together with Lemma 11.2 that

$$\|F(x^k)\| = \|F(x^k) - F(\bar{x})\| \leq L \cdot \|x^k - \bar{x}\|.$$

and

$$\|F(x^k) + F'(x^k)(x^{k+1} - x^k)\| \leq \eta_k \|F(x^k)\| \leq \eta_k L \|x^k - \bar{x}\| = o(\|x^k - \bar{x}\|).$$

Because of a), it follows that there exists some $\gamma \in (0, 1)$ and $l \geq 0$ such that for all $k \geq l$, it holds:

$$\|x^k - \bar{x}\| \leq \|x^{k+1} - x^k\| + \|x^{k+1} - \bar{x}\| \leq \|x^{k+1} - x^k\| + \gamma \|x^k - \bar{x}\|.$$

Hence, we get that

$$\|x^k - \bar{x}\| \leq \frac{1}{1-\gamma} \|x^{k+1} - x^k\| = \mathcal{O}(\|x^{k+1} - x^k\|).$$

It follows that

$$\|F(x^k) + F'(x^k)(x^{k+1} - x^k)\| \leq \eta_k L \|x^k - \bar{x}\| = o(\|x^k - \bar{x}\|) = o(\|x^{k+1} - x^k\|).$$

Hence, the statement follows with Theorem 11.3. ∎

Next, we consider the relationship between inexact Newton methods and Netwon-like methods. They can be identified with each other in the following sense:
First, we consider the Newton-like methods. They obtain the next direction s^k by solving $M_k s^k = -F(x^k)$. This equation can rewritten to

$$F'(x^k)s^k = -F(x^k) + r^k \quad \text{with } r^k = (F'(x^k) - M_k)s^k$$

Hence, s^k is an inexact solution of the Newton equation and

$$\|F(x^k) + F'(x^k)s^k\| = \|r^k\| = \|(F'(x^k) - M_k) \cdot s^k\|.$$

Now, we consider the inexact Newton method. Here, it holds that

$$F'(x^k)s^k = -F(x^k) + r^k.$$

We can choose a matrix M_k such that $M_k s^k = -F(x^k)$. For example,

$$M_k = F'(x^k) - \frac{r^k {s^k}^\top}{\|s^k\|^2}$$

is a valid choice. Then, the sequences $(x^k)_k$ and $(s^k)_k$ generated by the Newton-like method using M_k are identical with the sequences generated by the inexact Newton method. Hence, solving a system inexactly can be interpreted as solving the system exactly, but with a modified matrix.

13. Quasi-Newton Methods

Quasi-Newton methods also exist for nonlinear systems of equations. Here, we will only consider them for optimization problems of the form

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

with $f : \mathbb{R}^n \to \mathbb{R}$ and $f \in C^2(\mathbb{R}^n)$.
The idea of the Quasi-Newton methods is to use regular, symmetric matrices $H_k \in \mathbb{R}^{n \times n}$ as substitutes for the Hessian matrix. The matrices H_k shall fulfill for all $k \geq 0$ the so-called **Quasi-Newton equation (QNE)**

$$H_{k+1}(x^{k+1} - x^k) = \nabla f(x^{k+1}) - \nabla f(x^k). \tag{13.1}$$

The Hessian matrix $\nabla^2 f(x^{k+1})$ does, in general, not fulfill the Quasi-Newton equation.
Quasi-Newton methods have the following general structure:

Algorithm 13.1 General outline of the local Quasi-Newton method

1: Let $x^0 \in \mathbb{R}^n$, let $H_= \in \mathbb{R}^{n \times n}$ be a symmetric, regular matrix.
2: **for** $k = 0, 1, 2, \ldots$ **do**
3: **if** $\nabla f(x^k) = 0$ **then**
4: Stop
5: Compute $s^k \in \mathbb{R}^n$ by solving

$$H_k \cdot s^k = -\nabla f(x^k) \tag{13.2}$$

6: Set $x^{k+1} = x^k + s^k$
7: Compute symmetric, regular matrix $H_{k+1} := H(H_k, x^{k+1} - x^k, \nabla f(x^{k+1}) - \nabla f(x^k))$ by some updating formula satisfying the Quasi-Newton equation (13.1)

The motivation behind the use of the Quasi-Newton equation in algorithm 13.1 is that it satisfies the Dennis-Moré condition (see Corollary 11.4) under certain assumptions, i. e. superlinear convergence can be guaranteed under certain assumptions.
Intuitively, it makes sense that the Dennis-Moré condition is (roughly) fulfilled, because with the mean value theorem, we get that

$$H_{k+1}s^k = H_k(x^{k+1} - x^k) \overset{\text{(QNE)}}{=} \nabla f(x^{k+1}) - \nabla f(x^k) = \nabla^2 f(x^k + \xi)(x^{k+1} - x^k) \approx \nabla^2 f(x^k)s^k$$

More formally, we obtain the following superlinear convergence result:

Lemma 13.2 *Let $\bar{x} \in \mathbb{R}^n$ satisfy the second-order sufficient optimality conditions. If algorithm 13.1 generates the sequence $(x^k)_k$ with $x^k \to \bar{x}$ and*

$$\lim_{k\to} \|H_{k+1} - H_k\| = 0, \tag{13.3}$$

then H_k satisfies the Dennis-Moré condition and $(x^k)_k$ converges superlinearly to $\bar{x}$.

Proof. It follows that

$$\begin{aligned} \|(H_k - \nabla^2 f(x^k))s^k\| &\leq \|(H_k - H_{k+1})s^k\| + \|(H_{k+1} - \nabla^2 f(x^k))s^k\| \\ &\overset{\text{(QNE)}}{=} o(\|s^k\|) + \|\nabla f(x^{k+1}) - \nabla f(x^k) - \nabla^2 f(x^k)s^k\| \\ &\overset{\text{Taylor of } \nabla f}{=} o(\|s^k\|) \end{aligned}$$

Hence, the statement follows with Corollary 11.4. ∎

Because of (13.3), we will search for matrices H_{k+1} that are close to H_k.

Quasi-Newton Update Formulas

Our next goal is to generate a sequence of matrices $(H_k)_k$ according to the follwoing scheme: We start with a symmetric, regular matrix H_0 (and x^0) and compute x^1. Then, H_1 is computed by some particular update formula

$$H_1 = H(H_0, x^1 - x^0, \nabla f(x^1) - \nabla f(x^0)).$$

This procedure is continued. This means that if H_k is already computed, then H_{k+1} can be obtained by an update formula of the form

$$H_{k+1} = H(H_k, x^{k+1} - x^k, \nabla f(x^{k+1}) - \nabla f(x^k)).$$

The update formula is chosen in such a way that

1. H_{k+1} is again symmetric,
2. the Quasi-Newton equation (13.1) is satisfied,
3. the update effort is low and
4. the resulting Quasi-Newton method has a good (local) convergence.

In the following, we use the abbreviations

$$d^k := x^{k+1} - x^k \quad \text{and} \quad y^k := \nabla f(x^{k+1}) - \nabla f(x^k).$$

As a first type of update formulas, we consider the **Symmetric Rank 1-Formula (SR1)**

$$H_{k+1} = H_k + \gamma_k u^k {u^k}^\top,$$

where $\gamma_k \in \mathbb{R}$ and $u^k \in \mathbb{R}^n$ with $\|u^k\| = 1$ are "suitably chosen". For such an update formula, the Quasi-Newton equation has the form

$$H_{k+1}d^k = H_k d^k + \gamma_k ({u^k}^\top d^k) u^k \stackrel{!}{=} y^k. \tag{13.4}$$

If $y^k - H_k d^k = 0$, it follows that

$$\nabla f(x^{k+1}) = \nabla f(x^k) + y^k = \nabla f(x^k) + H_k d^k \stackrel{(s^k = d^k, \text{i. e. step size is 1})}{=} \nabla f(x^k) + H_k s^k \stackrel{(13.2)}{=} 0.$$

Hence, algorithm 13.1 terminates in this case and it is not necessary to compute H_{k+1}. Thus, let us assume that $y^k - H_k d^k \neq 0$:
From (13.4) and $\|u^k\| = 1$, it follows that

$$u^k = \pm \frac{y^k - H_k d^k}{\|y^k - H_k d^k\|}.$$

If $(y^k - H_k d^k)^\top d^k \neq 0$, it follows together with (13.4) that

$$\gamma_k {u^k}^\top d^k = \|y^k - H_k d^k\|$$

and

$$\gamma_k = \frac{\|y^k - H_k d^k\|^2}{(y^k - H_k d^k)^\top d^k}.$$

Hence, the only (SR1) formula where H_{k+1} is symmetric and fulfills (QNE) is

$$H_{k+1} = H_k + \frac{(y^k - H_k d^k)(y^k - H_k d^k)^\top}{(y^k - H_k d^k) \top d^k}.$$

However, this formula has several drawbacks:

- $(y^k - H_k d^k)^\top d^k$ can be zero.
- If $(y^k - H_k d^k)^\top d^k < 0$, then H_{k+1} may not be positive definite. In particular, H_{k+1} may not be regular.
- $s^k = -H_k^{-1} \nabla f(x^{k+1})$ may not be a descent direction.

As a new approach, we use the "next simplest" update formula: The **Symmetric Rank 2-Formula**:

$$H_{k+1} = H_k + \gamma_{k_1} u_1^k {u_1^k}^\top + \gamma_{k_2} u_2^k {u_2^k}^\top$$

In practice, the most successful Quasi-Newton methods use such Symmetric Rank 2-Formulas. In the following, we want to state the most important instances:

- **Broyden-Fletcher-Goldfarb-Shanno-Formula (BFGS update)**:

$$H_{k+1}^{\text{BFGS}} = H_k + \frac{y^k {y^k}^\top}{{y^k}^\top d^k} - \frac{H_k d^k (H_k d^k)^\top}{{d^k}^\top H_k d^k}$$

- **Davidon-Fletcher-Powell-Formula (DFP update)**:

$$H_{k+1}^{\mathrm{DFP}} = H_k + \frac{(y^k - H_k d^k){y^k}^\top + y^k(y^k - H_k d^k)^\top}{{y^k}^\top d^k} - \frac{(y^k - H_k d^k)^\top d^k}{({y^k}^\top d^k)^2} y^k {y^k}^\top$$

- **Broyden class**:

$$H_{k+1}^{\lambda} = (1-\lambda)\cdot H_{k+1}^{\mathrm{BFGS}} + \lambda \cdot H_{k+1}^{\mathrm{DFP}} = H_{k+1}^{\mathrm{BFGS}} + \lambda \cdot ({d^k}^\top H_k d^k) v^k {v^k}^\top, \quad \lambda \in \mathbb{R},$$

$$\text{where } v^k = \frac{y^k}{{y^k}^\top d^k} - \frac{H_k d^k}{{d^k}^\top H_k d^k}.$$

  This includes the special cases $H_{k+1}^0 = H_{k+1}^{\mathrm{BFGS}}$, $H_{k+1}^1 = H_{k+1}^{\mathrm{DFP}}$ and the (SR1) update for

$$\lambda = \frac{{y^k}^\top d^k}{{y^k}^\top d^k - {d^k}^\top H_k d^k}.$$

- **Convex Broyden class**: H_{k+1}^{λ} with $\lambda \in [0,1]$

In practice, the Quasi-Newton method based on the BFGS update has proven to be the most efficient one. The following Theorem shows that the BFGS update and the DFP update generate indeed symmetric matrices that fulfill (QNE). Regarding our approximability property (13.3) in Lemma 13.2, the Theorem also shows that, among all such matrices, the BFGS update and DFP update are "optimal" in a certain way:

Theorem 13.3 *Let H_k be symmetric and positive definite and let ${y^k}^\top d^k > 0$.*
Then, there exists a symmetric and positive definite matrix W with $W^2 d^k = y^k$.
For each such matrix W, it holds:

a) $H_+ = H_+^{DFP}$ solves the problem:

$$\begin{aligned} \min\ & \|W^{-1}(H_+ - H)W^{-1}\|_F \\ s.t.\ & H_+ = H_+^\top \\ & H_+ d^k = y^k \end{aligned}$$

b) $H_+ = H_+^{BFGS}$ solves the problem:

$$\begin{aligned} \min\ & \|W(H_+ - H)W\|_F \\ s.t.\ & H_+ = H_+^\top \\ & H_+ d^k = y^k, \end{aligned}$$

*where $\|\cdot\|_F$ denotes the **Frobenius norm**, i. e. $\|A\|_F := \sqrt{\sum_{i=1}^n \sum_{i=1}^n |a_{ij}|^2}$.*

The next Theorem shows that the convex Broyden class generates symmetric, positive definite matrices under certain assumptions:

Theorem 13.4

a) If ${y^k}^\top d^k \neq 0$ and ${d^k}^\top H_k d^k \neq 0$, then the matrices H_{k+1}^λ, $\lambda \in \mathbb{R}$ are symmetric and satisfy (QNE).

b) If H_k is positive definite and if ${y^k}^\top d^k > 0$, then the matrices H_{k+1}^λ with $\lambda \geq 0$ are positive definite.

Since the update formulas for the Broyden class are Rank 2-Formulas, it is also possible to determine Rank 2 Update-Formulas for H_k^{-1}. In the following Theorem, they are stated for the BFGS update and the DFP update:

Theorem 13.5 (Inverse update formulas) *Let H_k be positive definite and $B_k = H_k^{-1}$. Let ${y^k}^\top d^k > 0$. Let us denote $B_{k+1}^{BFGS} := (H_{k+1}^{BFGS})^{-1}$ and $B_{k+1}^{DFP} := (H_{k+1}^{DFP})^{-1}$.*
Then it holds:

a)

$$B_{k+1}^{BFGS} = B_k + \frac{(d^k - B_k y^k){d^k}^\top + d^k(d^k - B_k y^k)^\top}{{d^k}^\top y^k} - \frac{(d^k - B_k y^k)^\top y^k}{({d^k}^\top y^k)^2} d^k {d^k}^\top$$

b)

$$B_{k+1}^{DFP} = B_k + \frac{d^k {d^k}^\top}{{d^k}^\top y^k} - \frac{B_k y^k (B_k y^k)^\top}{{y^k}^\top B_k y^k}$$

Next, we present a local Quasi-Newton method with BFSG update that is using the inverse update formula and a cokonvergence theorem:

Algorithm 13.2 Local inverse BFGS method

1: Let $x^0 \in \mathbb{R}^n$ and let $B_0 \in \mathbb{R}^{n \times n}$ be a symmetric positive definite matrix
2: **for** $k = 0, 1, 2, \ldots$ **do**
3: $s^k := -B_k \cdot \nabla f(x^k)$
4: $x^{k+1} := x^k + s^k$
5: **if** $\nabla f(x^k) = 0$ **then**
6: Stop
7: $d^k := s^k$, $y^k := \nabla f(x^{k+1}) - \nabla f(x^k)$
8: **if** ${y^k}^\top \cdot d^k \leq 0$ **then**
9: Stop with error message
10: Compute

$$B_{k+1} = B_k + \frac{(d^k - B_k y^k){d^k}^\top + d^k(d^k - B_k y^k)^\top}{{d^k}^\top y^k} - \frac{(d^k - B_k y^k)^\top y^k}{({d^k}^\top y^k)^2} d^k {d^k}^\top$$

Theorem 13.6 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be two-times continuously differentiable with locally Lipschitz-continuous Hessian matrix $\nabla^2 f$. Let $\bar{x} \in \mathbb{R}^n$ fulfill the sufficient second-order optimality conditions.*
Then there exist some $\delta > 0$ and $\varepsilon > 0$ such that for any $x^0 \in B_\delta(\bar{x})$ and every symmetric positive definite start matrix $B_0 \in \mathbb{R}^{n \times n}$ with $\|B_0 - \nabla^2 f(\bar{x})^{-1}\| < \varepsilon$, algorithm 13.2 terminates with $x^k = \bar{x}$ or generates a sequence $(x^k)_k \subset B_\delta(\bar{x})$ that converges q-superlinearly to $\bar{x}$.

Algorithm 13.2 can be globalized similarly to the Newton's method when using Powell-Wolfe step length rule.

14. Trust-Region Methods

In this chapter, we consider again the unconstrained optimization problem

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

where $f : \mathbb{R}^n \to \mathbb{R}$ with $f \in C^2(\mathbb{R}^n)$.
The Trust Region method is again and an iterative method, i. e. we will compute points with an iteration of the form $x^{k+1} = x^k + s^k$. Its idea is to replace f by a simpler function, which is, locally, a good approximation of f. More precisely, we work with the quadratic model

$$q_k(s) = f_k + {g^k}^\top \cdot s + \frac{1}{2} s^\top H_k s,$$

where $f_k := f(x^k)$, $g^k := \nabla f(x^k)$ and $H_k := \nabla^2 f(x^k)$.
In other words, q_k is the Taylor polynomial with degree 2 of $f(x^k + s)$ around $s = 0$.
The function q_k is used as a predictor and substitute for f and s^k is obtained by a (restricted) minimization of q_k.
Taylors Theorem says that $f(x^k+s) = q_k(s)+o(\|s\|^2)$, i. e. $q_k(s)$ is a good approximation of f if $\|s\|$ is sufficiently small. Thus, we define a region, where we "trust" q_k to be a good model of f, the so-called **trust-region**

$$\{s \in \mathbb{R}^n : \|s\| \leq \Delta_k\}$$

The value Δ_k is called **trust-region radius**.
Then, the next direction s^k is obtained by solving the so-called **trust-region subproblem**

$$\begin{aligned} \min \quad & q_k(s) \\ \text{s. t.} \quad & s \in \mathbb{R}^n \\ & \|s\| \leq \Delta_k \end{aligned}$$

Next, we evaluate the quality of s^k. For this purpose we compare the **predicted reduction** of the model q_k

$$\text{pred}_k(s^k) := q_k(0) - q_k(s^k) = f_k - q_k(s^k)$$

with the **actual reduction** of f

$$\text{ared}_k(s^k) := f_k - f(x^k + s^k)$$

by computing the ratio

$$\rho_k(s^k) := \frac{\text{ared}_k(s^k)}{\text{pred}_k(s^k)}.$$

We choose a parameter $0 < \eta_1 < 1$ and decide on the quality of the model as follows:

- If $\rho_k(s^k) \leq \eta_1$, then the actual reduction is not satisfying compared to the predicted reduction. This means that the model was not good and the trust region was chosen too large. Hence, we reduce the trust region radius to $\Delta_{\min} \leq \Delta_{k+1} < \Delta_k$ and do not accept s^k. In this context, $\Delta_{\min} \geq 0$ is a parameter for the minimum radius that we allow in the whole method.

- If $\rho_k(s^k) > \eta_1$, then ared is good enough compared to pred and we accept s^k, i. e. we compute the next point $x^{k+1} = x^k + s^k$. Since the model is good on the trust region, we increase the trust region radius to $\Delta_{k+1} \geq \Delta_k > \Delta_{\min}$ for the next iteration $k+1$.

If $\|s^k\| < \Delta_k$, then s^k is lying in the topological inner of $\{s \in \mathbb{R}^n : \|s\| \leq \Delta_k\}$. Hence, s^k is the global minimum of q_k and s^k the Newton step in this case, i. e. $s^k = -\nabla^2 f(x^k)^{-1} \cdot \nabla f(x^k)$.
If $\nabla^2 f$ is not available or too difficult to compute, then an approximation H_k can be used (see chapter 13, e. g. BFGS, DFP, ...).
Sometimes, computing the exact solution of the trust region subproblem can be too time-consuming. But the Trust-Region method is already globally convergent, if the following condition is satisfied by the steps s^k:

Fraction of Cauchy Decrease: There exist constants $\alpha \in (0,1)$ and $\beta \geq 1$:

$$\|s^k\| \leq \beta\Delta_k, \quad \text{pred}_k(s^k) \geq \alpha \text{pred}_k(s_c^k), \tag{14.1}$$

where the so-called *Cauchy step* s_c^k is the unique solution of

$$\min\ q_k(s) \tag{14.2}$$
$$\text{s. t. } s = -tg^k \tag{14.3}$$
$$t \geq 0 \tag{14.4}$$
$$\|s\| \leq \Delta_k \tag{14.5}$$

(14.3) ensures that s_c^k is a positive multiple of the negative gradient. (14.5) guarantees that s_c^k is lying in the trust-region.
For proving the global convergence later on, we will need the following assumptions on f and H_k:

$$f \in C^2(\mathbb{R}^n) \text{ and bounded from below} \tag{14.6}$$
$$\text{There exists some constant } C_H > 0 \text{ such that it holds for all } k : \|H_k\| \leq C_H \tag{14.7}$$

The Trust-Region method has the following general form:

Algorithm 14.1 General structure of Trust-Region method

Choose parameter $\alpha \in (0, 1]$, $\beta \geq 1$, $0 < \eta_1 < \eta_2 < 1$, $0 < \gamma_0 < \gamma_1 < 1 < \gamma_2$ and $\Delta_{\min} \geq 0$
Choose starting point $x^0 \in \mathbb{R}^n$ and a trust-region radius $\Delta_0 > 0$ with $\Delta_0 \geq \Delta_{\min}$
for $k = 0, 1, 2, \ldots$ **do**
  if $g^k = 0$ **then**
    Stop and return x^k
  Choose a symmetric matrix $H_k \in \mathbb{R}^{n \times n}$
  Compute step s^k which satisfies the Fraction of Cauchy decrease condition (14.1)
  Compute $\rho_k(s^k)$
  if $\rho_k(s^k) > \eta_1$ **then**
    accept s^k and set $x^{k+1} = x^k + s^k$
  else
    discard s^k and set $x^{k+1} = x^k$
  update Δ_{k+1} (see Algorithm 14.2)

Algorithm 14.2 Update of Trust-Region Radius

if $\rho_k(s^k) \leq \eta_1$ **then**
  choose $\Delta_{k+1} \in [\gamma_0 \Delta_k, \gamma_1 \Delta_k]$
if $\rho_k(s^k) \in (\eta_1, \eta_2]$ **then**
  choose $\Delta_{k+1} \in [\max\{\Delta_{\min}, \gamma_1 \Delta_k\}, \max\{\Delta_{\min}, \Delta_k\}]$
if $\rho_k(s^k) > \eta_2$ **then**
  choose $\Delta_{k+1} \in [\max\{\Delta_{\min}, \Delta_k\}, \max\{\Delta_{\min}, \gamma_2 \Delta_k\}]$

In the following, we want to answer the questions:

a) Does Algorithm 14.1 converge globally?

b) How can we solve the trust-region subproblem?
How can we characterize its optimal solutions?

c) Does (a variant of) Algorithm 14.1 converge locally fast? What is its speed? Which conditions do we need to assume?

In the following, we will partition the steps s^k in two classes, the successful and the discarded steps:

Definition 14.3

1. the step s^k is called **successful** if $\rho_k(s^k) > \eta_1$
2. $S \subseteq \mathbb{N}_0$ denotes the set of all successful steps.

Next, we answer question a) about the global convergence:

Global convergence

In this section, we will show that Algorithm 14.1 globally converges under the assumptions (14.6) and (14.7). First, we show that the Fraction of Cauchy Decrease condtion (14.1) enables us the make the following estimation on the model decrease $\text{pred}(s^k)$:

Lemma 14.4 *Let the assumptions (14.6) and (14.7) be fulfilled. If $g^k \neq 0$ and if s^k satisfies the Fraction of Cauchy Decrease Condition (14.1), then*

$$pred_k(s^k) \geq \frac{\alpha}{2}\|g^k\| \min\{\Delta_k, \|g^k\|/C_H\}$$

Proof. It is $\text{pred}_k(s_c^k) = \varphi(\tau^*)$, where τ^* is the maximum of the function

$$\varphi(\tau) = \text{pred}_k(-\tau g^k) = q_k(0) - q_k(-\tau g^k) = \tau\|g^k\|^2 - \frac{1}{2}\tau^2 {g^k}^\top H_k g^k$$

on $[0, \kappa]$ with $\kappa = \frac{\Delta_k}{\|g^k\|}$.
We distinguish two cases:

1. If ${g^k}^\top H_k g^k \leq 0$, then both summands of φ become larger for increasing τ, i. e. $\tau^* = \kappa$ and

$$\varphi(\tau^*) = \kappa\|g^k\|^2 - \frac{1}{2}\underbrace{\kappa^2 {g^k}^\top H_k g^k}_{\leq 0} \geq \kappa\|g^k\|^2 = \|g^k\|\Delta_k.$$

2. If ${g^k}^\top H_k g^k > 0$, then $\varphi'(\tau) = \|g^k\|^2 - \tau {g^k}^\top H_k g^k$ and $\varphi''(\tau) = -{g^k}^\top H_k g^k < 0$. Hence, φ is strictly concave (i. e. $-\varphi$ is strictly convex) in that case and the stationary point $\tau = \frac{\|g^k\|^2}{{g^k}^\top H_k g^k}$ is the global maximum if it lies in $[0, \kappa]$. Hence, there are two possibilities in that case:

   a) $\tau^* = \kappa \leq \frac{\|g^k\|^2}{{g^k}^\top H_k g^k}$ and

$$\varphi(\tau^*) = \kappa\|g^k\|^2 - \frac{1}{2}\kappa^2 {g^k}^\top H_k g^k \geq \frac{1}{2}\kappa\|g^k\|^2 = \frac{1}{2}\|g^k\|\Delta_k$$

   b) $\tau^* = \frac{\|g^k\|^2}{{g^k}^\top H_k g^k} < k$ and

$$\varphi(\tau^*) = \frac{1}{2}\frac{\|g^k\|^4}{{g^k}^\top H_k g^k} = \frac{1}{2}\frac{\|g^k\|^4}{|{g^k}^\top H_k g^k|} \overset{(*)}{\geq} \frac{1}{2}\frac{\|g^k\|^2}{\|H_k\|} \overset{(14.7)}{\geq} \frac{1}{2}\frac{\|g^k\|^2}{C_H}$$

   ($(\star)$ follows from Cauchy-Schwarz inequality and $\|Ax\| \leq \|A\| \cdot \|x\|$) ∎

Next, we want to show that the search for a successful step is always successful. For this purpose, we need the next Lemma. If we consider it for a non-stationary point $x = x^k$, it says that if the trust radius is sufficiently small, any step size s^k fulfilling (14.1) surpasses any fixed quality threshold $\rho_k(s^k) > \eta_1$, no matter how close to 1 we choose $\eta_1 \in (0, 1)$.

Lemma 14.5 *Let $f \in C^1(\mathbb{R})$, let $x \in \mathbb{R}^n$ with $\nabla f(x) \neq 0$. Then, for every $\eta \in (0,1)$, there exist some $\delta = \delta(x,\eta)$ and $\Delta = \Delta(x,\eta) > 0$ such that*

$$\rho_k(s^k) > \eta$$

for all $\|x^k - x\| \leq \delta$ and all $\Delta_k \in (0, \Delta]$, $s^k \in \mathbb{R}^n$ and $H_0 = H_0^\top \in \mathbb{R}^{n\times n}$ which fulfill assumption (14.7) and the Fraction of Cauchy Drecrease condition (14.1).

Proof. Due to the mean value theorem, there is a $\tau \in [0,1]$ such that

$$\begin{aligned}
\operatorname{pred}_k(s^k) - \operatorname{ared}_k(s^k) &= f(x^k + s^k) - f_k + \underbrace{q_k(0)}_{=f_k} - q_k(s^k) \\
&= f(x^k + s^k) - f_k - {g^k}^\top s^k - \frac{1}{2} {s^k}^\top H_k s^k \\
&= \nabla f(x^k + \tau s^k)^\top s^k - {g^k}^\top s^k - \frac{1}{2} {s^k}^\top H_k s^k \\
&\overset{(14.7),(14.1),\text{Cauchy-Schwarz-}}{\leq} \beta \|\nabla f(x^k + \tau s^k) - g^k\| \Delta_k + \frac{\beta^2 C_H}{2} \Delta_k^2.
\end{aligned}$$

With (14.1) and Lemma 14.4, it also follows that

$$\operatorname{pred}_k(s^k) \geq \frac{\alpha}{2} \|g^k\| \min\{\Delta_k, \|g^k\|/C_H\}.$$

Since ∇f is continuous, we can choose a sufficiently small $\delta > 0$ such that $\|g^k\| \geq \varepsilon := \frac{\|\nabla f(x)\|}{2}$ for all x^k with $\|x^k - x\| \leq \delta$.
For all $0 < \Delta \leq \frac{\varepsilon}{C_H}$ and all $\Delta_k < \Delta$, we obtain

$$\operatorname{pred}_k(s^k) \geq \frac{\alpha}{2} \|g^k\| \Delta_k \geq \frac{\alpha}{2} \varepsilon \Delta_k. \tag{14.8}$$

Because of

$$\|x^k - x\| \leq \delta, \quad \|x^k + \tau s^k - x\| \leq \|x^k - x\| + \|s^k\| \overset{(14.1)}{\leq} \delta + \beta\Delta,$$

it holds for $\delta + \Delta \to 0$ that

$$g^k = \nabla f(x^k) \to \nabla f(x) \quad \text{and} \quad \nabla f(x^k + \tau s^k) \to \nabla f(x).$$

Hence, if δ and Δ are chosen sufficiently small, it holds that

$$\operatorname{pred}_k(s^k) - \operatorname{ared}_k(s^k) \leq \beta \|\nabla f(x^k + \tau s^k) - g^k\| \Delta_k + \frac{\beta^2 C_H}{2} \Delta_k^2 < (1-\eta) \frac{\alpha}{2} \varepsilon \Delta_k. \tag{14.9}$$

We then get that

$$\rho_k(s^k) = \frac{\operatorname{ared}_k(s^k)}{\operatorname{pred}_k(s^k)} = 1 - \frac{\operatorname{pred}_k(s^k) - \operatorname{ared}_k(s^k)}{\operatorname{pred}_k(s^k)} \overset{(14.8),(14.9)}{>} 1 - \frac{(1-\eta)\frac{\alpha}{2}\varepsilon\Delta_k}{\frac{\alpha}{2}\varepsilon\Delta_k} = \eta. \quad \blacksquare$$

We get the following Corollary immediately:

Corollary 14.6 *If Algorithm 14.1 does not terminate after a finite number of iterations and if the assumptions (14.6) and (14.7) hold, then Algorithm 14.1 generates infinitely many successful steps.*

Proof. The proof is done by contradiction. Assume the claim is false. This means that there exists some $l \geq 0$ with

$$x^k = x^l \quad \text{and} \quad \rho_k(s^k) \leq \eta_1 \quad \forall k \geq l. \tag{14.10}$$

Hence, it holds that

$$\Delta_k \leq \gamma_1 \Delta_{k-1} \leq \cdots \leq \gamma_1^{k-l} \Delta_l \overset{k\to\infty}{\longrightarrow} 0$$

Since $g^l \neq 0$, Lemma 14.5 can be applied with $x = x^l$ and $\eta = \eta_1$.
We then get some $\Delta > 0$ such that $\rho_k(s^k) > \eta_1$ for all $k \geq l$ with $\Delta_k \leq \Delta$.
Since $\Delta_k \to 0$, there exists a smallest such k. Then step s^k would be successful, which is a contradiction to (14.10). ∎

We need one more Lemma for proving the global convergence:

Lemma 14.7 *Consider Algorithm 14.1 with the assumptions (14.6) and (14.7). If $K \subseteq S$ is an infinite set with $\|g^k\| \geq \varepsilon > 0$ for all $k \in K$, then it holds that*

$$\sum_{k \in K} \Delta_k < \infty.$$

Proof. For all $k \in K \subseteq S$, step s^k is successful and thus, it holds that

$$\begin{aligned} f(x^k) - f(x^{k+1}) &= \operatorname{ared}_k(s^k) \\ &> \eta_1 \operatorname{pred}_k(s^k) \\ &\overset{\text{Lemma 14.4}}{\geq} \eta_1 \frac{\alpha}{2} \|g^k\| \min\{\Delta_k, \frac{\|g^k\|}{C_H}\} \\ &\geq \eta_1 \frac{\alpha}{2} \varepsilon \min\{\Delta_k, \frac{\varepsilon}{C_H}\} \end{aligned}$$

Since $f(x^k) \geq f(x^{k+1})$ for all k, it follows that

$$\begin{aligned} f(x^0) - f(x^k) &= \sum_{l \in S, l<k} (f(x^l) - f(x^{l+1})) \\ &\geq \sum_{l \in K, l<k} (f(x^l) - f(x^{l+1})) \\ &\geq \eta_1 \frac{\alpha}{2} \varepsilon \sum_{l \in K, l<k} \min\{\Delta_l, \frac{\varepsilon}{C_H}\} =: S_k \end{aligned}$$

If $\sum_{l \in K} \Delta_l = \infty$, then the sequence $(S_k)_k$ would be unbounded and we would get that $f(x^k) \to -\infty$, which is a contradiction to assumption (14.6). ∎

Now, we can show the global convergence of the Trust-Region method:

Theorem 14.8 *Under the assumptions (14.6) and* (14.7)*, algorithm 14.1 either terminates with a stationary point x^k or it generates an infinite sequence $(x^k)_k$ with*

$$\liminf_{k\to\infty} \|g^k\| = 0.$$

If ∇f is uniformaly continuous on $\Omega \subseteq \mathbb{R}^n$ with $(x^k)_k \subseteq \Omega$, then

$$\lim_{k\to\infty} \|g^k\| = 0.$$

Proof. We only show the first part of the theorem.
It follows from Corollary 14.6 that Algorithm 14.1 terminates or generates infinitely many successful steps.
Let us assume that $\liminf_{k\to\infty} \|g^k\| \neq 0$.
Then there exists some $\varepsilon > 0$ such that $\|g^k\| \geq \varepsilon$ for all $k \geq 0$.
With Lemma 14.7, it follows that

$$\sum_{k\in S} \Delta_k < \infty.$$

Hence, it holds that $\Delta_k \to 0$ for $S \ni k \to \infty$.
Additionally, it holds for all $k > l$ that

$$\|x^k - x^l\| \leq \sum_{i\in S, l\leq i<k} \|s^i\| \leq \beta \sum_{i\in S, l\leq i<k} \Delta_i \leq \beta \sum_{i\in S, i\geq l} \Delta_i \overset{l\to\infty}{\to} 0.$$

Hence, $(x^k)_k$ is a Cauchy sequence and $x^k \to \bar{x}$ with $\bar{x} \in \mathbb{R}^n$.
With the continuity of ∇f, it follows that $\|\nabla f(\bar{x})\| \geq \varepsilon$.
The application of Lemma 14.5 with $x = \bar{x}$ and $\eta = \eta_2$ yields to constants $L > 0$ and $\Delta > 0$ such that $p_k(s^k) > \eta_2$ for all $k \geq L$ with $\Delta_k \leq \Delta$.
Next, we show inductively that

$$\Delta_k \geq \min\{\Delta_L, \gamma_0\Delta\} \quad \forall k \geq L \tag{14.11}$$

If $k = L$, then (14.11) holds trivially.
Suppose (14.11) holds for some arbitrary but fixed k.
If $\Delta_k > \Delta$, then $\Delta_{k+1} \geq \gamma_0\Delta_k > \gamma_0\Delta$.
If $\Delta_k \leq \Delta$, then $\rho_k(s^k) > \eta_2$, and, thus, $\Delta_{k+1} \geq \Delta_k \geq \min\{\Delta_L, \gamma_0\Delta\}$.
Hence, (14.11) is true, which is a contradiction to $\lim_{S\ni k\to\infty} \Delta_k = 0$. ∎

Next, we state another converge result that requires a choice of $\Delta_{\min} > 0$:

Theorem 14.9 *Let the assumptions (14.6) and (14.7) be fulfilled and let $\Delta_{min} > 0$. Then, Algorithm 14.1 either terminates with a stationary point or it generates a sequence $(x^k)_k$ whose accumulation points are stationary.*

Proof. Consider the case that Algorithm 14.1 does not terminate.
Let $\bar{x}$ be an accumulation point and $K \subseteq S$ such that $(x^k)_K \to \bar{x}$.
Let us assume that $\nabla f(\bar{x}) \neq 0$.
Then there exist $l > 0$ and $\varepsilon > 0$ with $\|g^k\| \geq \varepsilon$ for all $k \in K$ with $k \geq l$.
Hence, it follows by Lemma 14.7 that

$$\sum_{k \in K} \Delta_k < \infty. \tag{14.12}$$

By Lemma 14.5, there exist $\delta > 0$ and $\Delta > 0$ with $k \in S$, if $x^k \in \overline{B_\delta(\bar{x})}$ and $\Delta_k < \Delta$.
Let $k \in K$ be sufficiently large. Then it holds that $x^k \in \overline{B_\delta(\bar{x})}$.
If $k-1 \in S$, then it follows that $\Delta_k \geq \Delta_{\min} > 0$.
If $k-1 \notin S$, then $x^{k-1} = x^k$ and $\Delta_{k-1} > \Delta$ (otherwise, s^{k-1} would be successful). It follows that

$$\Delta_k \geq \gamma_0 \Delta_{k-1} > \gamma_0 \Delta.$$

Hence, it holds for large $k \in K$ that $\Delta_k \geq \min\{\gamma_0 \Delta, \Delta_{\min}\}$, which is a contradiction to (14.12). ∎

Characterization of optimal solutions of the subproblem

Next, we want to characterize the optimal solutions of the Trust-Region subproblem:

$$\begin{aligned} \min \quad & q_k(s) \\ \text{s. t.} \quad & s \in \mathbb{R}^n \\ & \|s\| \leq \Delta_k \end{aligned} \tag{14.13}$$

Theorem 14.10

a) Problem (14.13) has at least one (globally) optimal solution.

b) $s^k \in \mathbb{R}^n$ is an optimal solution of (14.13) if and only if there exists a $\lambda \geq 0$ such that

$$\|s^k\| \leq \Delta_k \tag{14.14}$$

$$\lambda \geq 0, \quad \lambda(\|s^k\| - \Delta_k) = 0 \tag{14.15}$$

$$(H_k + \lambda I)s^k = -g^k \tag{14.16}$$

$$H_k + \lambda I \text{ is positive semidefinite} \tag{14.17}$$

c) If (14.14)-(14.16) holds and if $H_k + \lambda I$ is positive definite, then s^k is the unique optimal solution of (14.13).

Proof.

a) Since the feasible set of (14.13) is compact and q_k is continuous, it follows that (14.13) has an optimal solution.

b) "⇒":
Let s^k be an optimal solution of (14.13).
Let $y^k := \nabla q_k(s^k) = H_k s^k + g^k$.
It is clear that (14.14) is satisfied.
For proving (14.15) and (14.16), we distinct two cases:

Case 1: $\|s^k\| < \Delta_k$
in this case, (14.15) and (14.16) hold true with $\lambda = 0$ since s^k is a locally optimal solution and thus satisfies the first-order necessary optimality condition, i. e.

$$\nabla q_k(s^k) = 0 \Leftrightarrow H_k s^k + g^k = 0 \Leftrightarrow (H_k + 0 \cdot I)s^k = -g^k.$$

Case 2: $\|s^k\| = \Delta_k$
Let us assume there exists no $\lambda \geq 0$ such that (14.15) and (14.16) hold true.
Then, it follows that $y^k \neq 0$ and that $\alpha_k = \sphericalangle(y^k, s^k) \neq \pi$.
Hence, it holds that

$$\cos(\alpha_k) = \frac{{y^k}^\top s^k}{\|y^k\| \cdot \|s^k\|} > -1.$$

We define $v^k = -\frac{y^k}{\|y^k\|} - \frac{s^k}{\|s^k\|}$.
It follows that

$${y^k}^\top v^k = -\frac{{y^k}^\top y^k}{\|y^k\|} - \frac{{y^k}^\top s^k}{\|s^k\|} = -\|y^k\|(1 + \cos(\alpha_k)) < 0.$$

Hence v^k is a descent direction of q_k in s^k.
Moreover, it follows that

$$\begin{aligned}\left[\frac{d}{dt}\frac{1}{2}\|s^k + tv^k\|^2\right]_{t=0} &= {v^k}^\top s^k \\ &= -\frac{{y^k}^\top s^k}{\|y^k\|} - \frac{{s^k}^\top s^k}{\|s^k\|} \\ &= -\left(\frac{{y^k}^\top s^k}{\|y^k\|} + \|s^k\|\right) \\ &= -\|s^k\|(\cos(\alpha_k) + 1) < 0.\end{aligned}$$

Hence, it holds that $\|s^k + tv^k\| < \|s^k\| \leq \Delta_k$ for small $t > 0$. Since v^k is a descent direction, this contradicts the optimality of s^k.
Hence, the assumption was wrong and (14.15) and (14.16) hold true.

For (14.17), it is sufficient to show that

$$d^\top (H_k + \lambda I)d \geq 0 \quad \forall d \in \mathbb{R}^n \text{ with } d^\top s^k < 0,$$

because the sign of d does not matter and the case $d^\top s^k = 0$ follows due to continuity (since if $d^\top s^k = 0$ and $d^\top (H_k + \lambda I)d < 0$, then $(d - ts^k)^\top (H_k + \lambda I)(d -$

$ts^k) < 0$ for small t and it is $(d - ts^k)^\top s^k = \underbrace{d^\top s^k}_{=0} - t\|s^k\|^2 < 0$.)

Let $d \in \mathbb{R}^n$ with $d^\top s^k < 0$. We define $t := -\frac{2d^\top s^k}{\|d\|^2} > 0$.
Then, it follows that

$$\|s^k + td\|^2 = \|s^k\|^2 + 2ts^{k\top} d + t^2\|d\|^2 = \|s^k\|^2 \leq \Delta_k^2$$

and we get that

$$\begin{aligned} 0 \leq q_k(s^k + td) - q_k(s^k) &= tg^k d + ts^{k\top} H_k d + \frac{t^2}{2} d^\top H_k d \\ &= ty^{k\top} d + \frac{t^2}{2} d^\top H_k d \\ &\overset{(14.16)}{=} -t\lambda s^{k\top} d + \frac{t^2}{2} d^\top H_k d \\ &\overset{\text{def. } t}{=} \frac{t^2}{2}\lambda\|d\|^2 + \frac{t^2}{2} d^\top H_k d \\ &= \frac{t^2}{2} d^\top (H_k + \lambda I) d. \end{aligned}$$

Hence, (14.17) follows.

"$\Leftarrow$":
Let (14.14)-(14.17) be fulfilled for some $\lambda \geq 0$ and let $s^k \in \mathbb{R}^n$.
Let $h \in \mathbb{R}^n$ be arbitrary with $\|h\| \leq \Delta_k$.
Let $d := h - s^k$ and $y^k := H_k s^k + g^k$.
Then it follows that

$$\begin{aligned} q_k(h) - q_k(s^k) &= y^{k\top} d + \frac{1}{2} d^\top H_k d \\ &\overset{(14.16)}{=} -\lambda s^{k\top} d + \frac{1}{2} d^\top H_k d \\ &\overset{(14.17)}{\geq} -\lambda s^{k\top} d - \frac{1}{2}\lambda\|d\|^2 \\ &= -\frac{\lambda}{2}(2s^{k\top} d + \|d\|^2) \\ &= -\frac{\lambda}{2}(\|h\|^2 - \|s^k\|^2) \\ &\overset{(\star)}{\geq} 0 \end{aligned}$$

If $\lambda = 0$, then $(\star)$ is clear.

If $\lambda > 0$, then $(\star)$ follows from $\|h\| \leq \Delta_k \overset{(14.15)}{=} \|s^k\|$.
Hence, s^k solves (14.13).

c) Let $H_k + \lambda I$ be positive definite. Then it follows for $h \neq s^k$ with $\|h\| \leq \Delta_k$, analogously to the previous steps, that

$$q_k(h) - q_k(s^k) = \cdots = -\lambda s^{k\top} d + \frac{1}{2} d^\top H_k d > -\lambda s^{k\top} d - \frac{1}{2}\|d\|^2 = \cdots \geq 0.$$

This show the uniqueness of the optimal solution s^k. ∎

Fast local convergence

Next, we briefly consider under which assumptions we can expect fast local convergence. For this purpose we use the exact Hessian matrix $H_k = \nabla^2 f(x^k)$.
Additionally, if the Newton step $s_n^k = -H_k^{-1} g^k$ exists and it fulfills the Fraction of Cauchy Decrease condition, then we chosse $s^k := s_n^k$. Otherwise, we compute a step which satisfies the Fraction of Cauchy Decrease condition.

Then, we get the following convergence result:

Theorem 14.11 *Let $f \in C^2(\mathbb{R}^n)$, let $N_0 = \{x \in \mathbb{R}^n : f(x) \leq f(x^0)\}$ be compact. Suppose Algorithm 14.1 generates a sequence which has an accumulation point $\bar{x}$ such that $\nabla^2 f(\bar{x})$ is positive definite. Then it holds:*

a) $x^k \to \bar{x}$, $g^k \to 0$ for $k \to \infty$ and $\nabla f(\bar{x}) = 0$.

b) There exists an $l \geq 0$ such that for all $k \geq l$, it holds that every step s^k is successful and that $\Delta_k \geq \Delta_l > 0$.

c) There exists an $l' \geq l$ sucht that $s^k = s_n^k$ for all $k \geq l'$, i. e. Algorithm 14.1 turns into Newton's method after iteration l' and, thus, the convergence results of Newton's method apply.

Part III.

Constrained Optimization

15. Basics and Optimality Conditions for Constrained Nonlinear Programs

In this and the following chapters, we consider **constrained Nonlinear Programs**

$$\begin{aligned} \min\ & f(x) \\ \text{s.t.}\ & g(x) \leq 0 \\ & h(x) = 0, \end{aligned} \tag{15.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$ and $h : \mathbb{R}^n \to \mathbb{R}^p$ are continuously differentiable. The set

$$X = \{x \in \mathbb{R}^n : g(x) \leq 0, h(x) = 0\}$$

is called the **feasible set** of (15.1).
We define the index sets $U := \{1, \ldots, m\}$ and $G := \{1, \ldots, p\}$.

Definition 15.1

1. $x \in \mathbb{R}^n$ is called feasible if $x \in X$.
2. $\mathcal{A}(x) = \{i : 1 \leq i \leq m, g_i(x) = 0\}$ denotes the index set of **active inequalities**.
3. $\mathcal{I}(x) = \{i : 1 \leq i \leq m, g_i(x) < 0\}$ denotes the index set of **inactive inequalities**.

Definition 15.2 For $v \in \mathbb{R}^n$ and $J \subseteq \{1, \ldots, n\}$, we denote by $v_J \in \mathbb{R}^{|J|}$ the vector composed of the components v_j, $j \in J$.

Example 15.3 Let $g(x) = x_1^2 + x_2^2 - 1$. Then $X = \{x \in \mathbb{R}^2 : g(x) \leq 0\}$ is the circle (including the inner part) around $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ with radius 1. It holds that $\nabla g(x) = 2x$. The gradient $\nabla g(x)$ is orthogonal to the tangent plane of the circle in x (pointing out of X). For example, for $\bar{x} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, we get $\nabla g(\bar{x}) = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$.
The situation is illustrated in Figure 15.1.

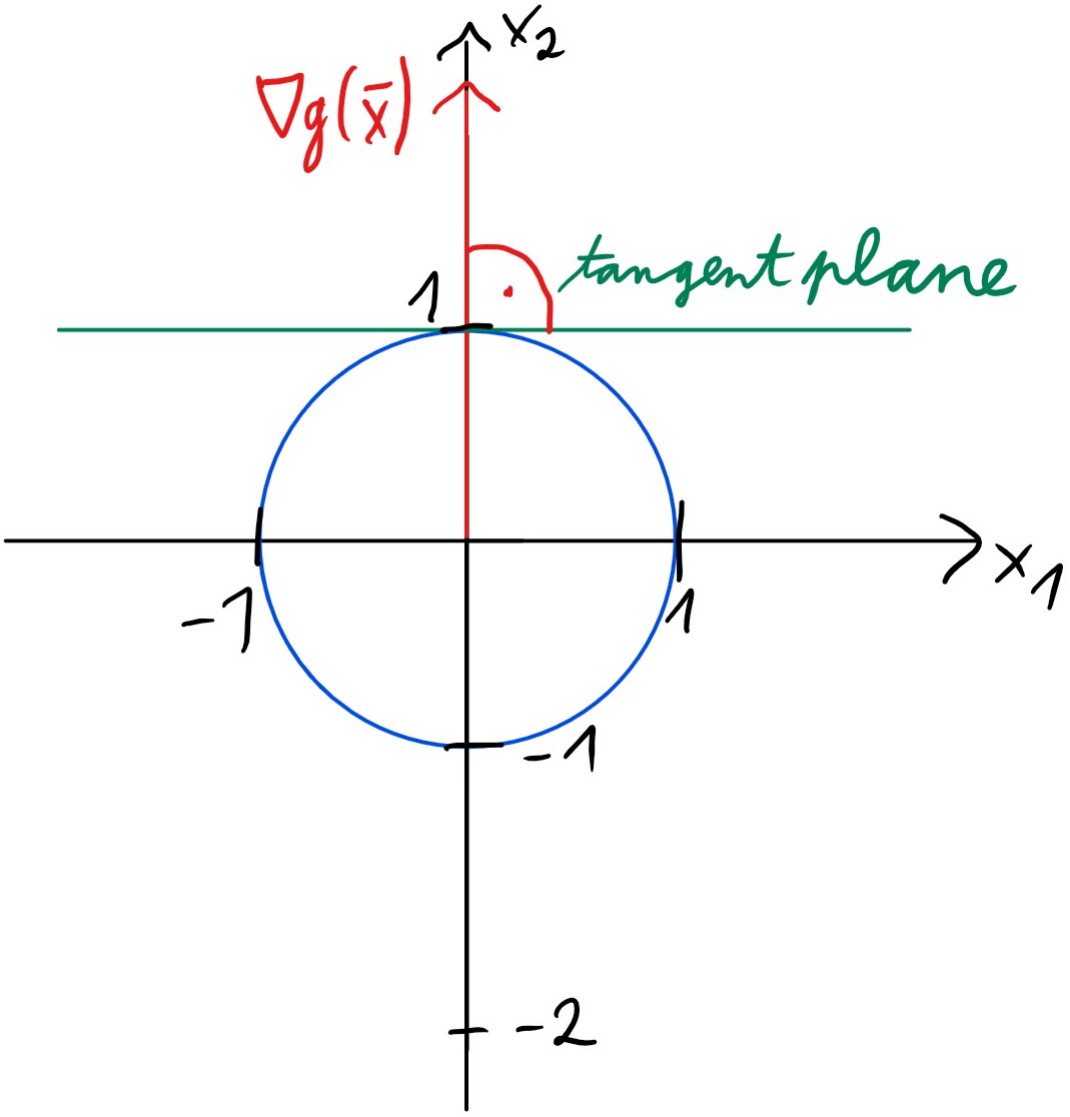


Figure 15.1.: Illustration of Example 15.3

Next, we derive optimality conditions for constrained Nonlinear Programs.

First-Order Necessary Optimality Conditions

For the first-order optimality conditions, we need to define special cones:

Definition 15.4

a) $K \subseteq \mathbb{R}^n$ is called **cone** if $\lambda x \in K$ for all $\lambda > 0$, $x \in K$

b) Let $\emptyset \neq M \subseteq \mathbb{R}^n$. Then

$$T(M,x) := \{d \in \mathbb{R}^n : \exists \eta_k > 0, x^k \in M : \lim_{k\to\infty} x^k = x, \lim_{k\to\infty} \eta_k(x^k - x) = d\}$$

is called **tangential cone** of M in $x \in M$.

Example 15.5

1. Informally, the tangential cone $T(M,x)$ is the set of all tangents to M at x.

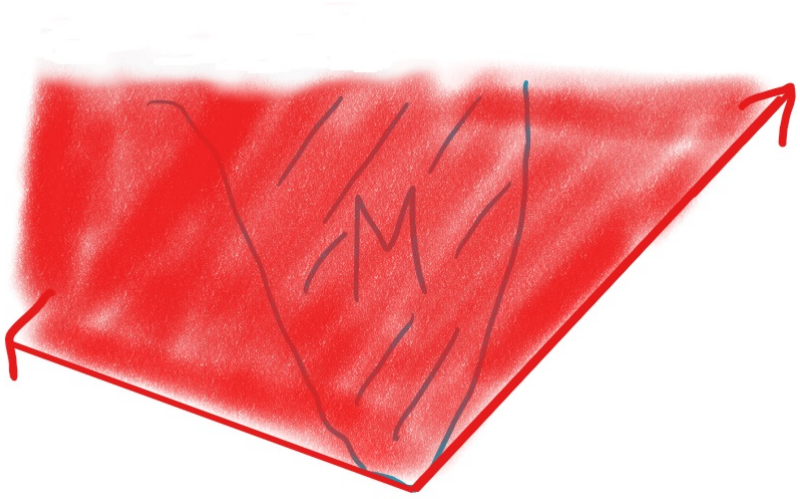


Figure 15.2.: Illustration of a tangential cone

2. One can show that $T(M,x)$ is a cone.

3. Let us focus on unit vectors contained in $T(M,x)$, i. e. we take $\frac{1}{\eta_k} := \|x^k - x\|$ in Definition 15.4b).
It holds that a unit vector $u \in T(M,x) \Leftrightarrow \exists (x^k)_k \subset M \to x : \lim_{k\to\infty} \frac{x^k - x}{\|x^k - x\|} = d$.
Let $M = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in \mathbb{R}^3 : x_1^2 + x_2^2 + x_3^2 = 4 \right\}$ be the sphere around $\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$ with radius 2.
We want to determine the tangential cone at $x = \begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix}$:
Consider $(x^k)_k \subset M \to x$, i. e.

$$\begin{pmatrix} x_1^k \\ x_2^k \\ \sqrt{4 - (x_1^k)^2 - (x_2^k)^2} \end{pmatrix} \to \begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix}.$$

Then it holds that

$$x^k - x = \begin{pmatrix} x_1^k \\ x_2^k \\ \sqrt{4-(x_1^k)^2-(x_2^k)^2}-2 \end{pmatrix} = \begin{pmatrix} x_1^k \\ x_2^k \\ \frac{-(x_1^k)^2-(x_2^k)^2}{\sqrt{4-(x_1^k)^2-(x_2^k)^2}+2} \end{pmatrix}$$

We note that $\|x^k - x\| \approx \sqrt{(x_1^k)^2 + (x_2^k)^2}$, because the third component gets small quickly as $x^k \to x$ and can therefore be dropped from consideration.
Thus, we get that

$$\frac{x^k - x}{\|x^k - x\|} \approx \frac{1}{\sqrt{(x_1^k)^2 + (x_2^k)^2}} \begin{pmatrix} x_1^k \\ x_2^k \\ 0 \end{pmatrix}.$$

Hence, $T(M, x)$ is here the plane where $x_3 = 0$.

Now, we can state the first necessary optimality condition:

Theorem 15.6 (Necessary Optimality Condition) *Let $\bar{x} \in \mathbb{R}^n$ be a local minimum of (15.1). Then it holds that*

a) $\bar{x} \in X$

b) $\nabla f(\bar{x})^\top d \geq 0$ *for all* $d \in T(X, \bar{x})$

Proof.

a) trivial, this is the feasibility of $\bar{x}$.

b) Let $(x^k)_k \subseteq X$ with $x^k \to x$, $\eta_k > 0$, $d^k := \eta_k(x^k - x) \to d$.
Since $\bar{x}$ is a local minimum, it follows that $f(x^k) - f(\bar{x}) \geq 0$ for large k.
With Taylor's theorem, we can conclude that

$$\begin{aligned} 0 &\leq \eta_k(f(x^k) - f(\bar{x})) \\ &= \eta_k \nabla f(\bar{x})^\top (x^k - \bar{x}) + \eta_k o(\|x^k - \bar{x}\|) \\ &= \nabla f(\bar{x})^\top d^k + \|d^k\| \frac{o(\|x^k - \bar{x}\|)}{\|x^k - \bar{x}\|} \overset{k\to\infty}{\longrightarrow} \nabla f(\bar{x})^\top d. \end{aligned}$$

The set $T(M, x)$ is a bit bulky, i. e. the condition in Theorem 15.6 b) is difficult to check. Hence, we want to want to find a more applicable condtion. Our idea is to linearize the constraints and approximate X locally. For the linearization, we use Taylor's Theorem.

Definition 15.7

$$T_l(g, h, x) := \{d \in \mathbb{R}^n : \nabla g_i(x)^\top d \leq 0, i \in \mathcal{A}(x), \nabla h(x)^\top d = 0\}$$

is called **linearized tangential cone** at $x \in X$.

Remark 15.8

a) Inactive constraints are locally not important, thus they are neglected in Definition 15.7.

b) Linearizing all constraints in $\bar{x}$ leads to

$$X_l(\bar{x}) = \{x : g(\bar{x}) + \nabla g(\bar{x})^\top (x - \bar{x}) \leq 0, h(\bar{x}) + \nabla h(\bar{x})^\top (x - \bar{x}) = 0\}$$

One can show that $T(X_l(\bar{x}), \bar{x}) = T_l(g, h, \bar{x})$.

c) Checking if $d \in T(X, \bar{x})$ is typically difficult.
Checking if $d \in T_l(g, h, \bar{x})$ is easy.
Hence, we would like to substitute $T(X, \bar{x})$ in Theorem 15.6 by $T_l(g, h, \bar{x})$.

$T(X, x)$ and $T_l(g, h, x)$ have the following relation:

Lemma 15.9 *For all $x \in X$, it holds: $T(X, x) \subseteq T_l(g, h, x)$.*

Proof. Let $d = \lim_{k\to\infty} d^k$ with $d^k = \eta_k(x^k - x)$, $\eta_k > 0$ and $(x^k)_k \subset X$ with $x^k \to x$. Then it holds:

$$0 \geq \eta_k(g_i(x^k) - g_i(x)) = \nabla g_i(x)^\top d^k + \eta_k o(\|x^k - x\|), \quad i \in \mathcal{A}(x)$$

$$0 = \eta_k(h_i(x^k) - h_i(x)) = \nabla h_i(x)^\top d^k + \eta_k o(\|x^k - x\|), \quad i \in G$$

For $k \to \infty$, this implies

$$\nabla g_i(x)^\top d \leq 0, \quad i \in A(x)$$

$$\nabla h_i(x)^\top d = 0, \quad i \in G$$

Thus, if $d \in T(X, x)$, then $d \in T_l(g, h, x)$ ∎

Example 15.10 In general, the other inclusion "$\supseteq$" does not hold in Lemma 15.9. Moreover, $T(X, x)$ depends on X only, while $T_l(g, hx)$ depends on the representation of $X = \{x \in \mathbb{R}^n : g(x) \leq 0, h(x) = 0\}$, which is not unique:
Let

$$X = \{x \in \mathbb{R}^2 : -x_1 - 1 \leq 0, x_1 - 1 \leq 0, x_2 = 0\},$$

i. e.

$$g(x) = \begin{pmatrix} -x_1 - 1 \\ x_1 - 1 \end{pmatrix} \quad \text{and} \quad h(x) = x_2.$$

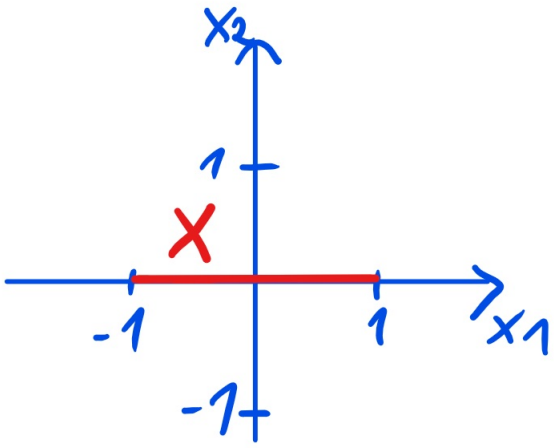


Figure 15.3.: Feasible set X

We consider the point $\bar{x} = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$. For this point, it holds that $\mathcal{A}(\bar{x}) = \{1\}$, $\nabla g_1(\bar{x}) = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$ and $\nabla h(\bar{x}) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$.
Hence, we get that

$$T_l(g,h,x) = \left\{ d \in \mathbb{R}^2 : \begin{pmatrix} -1 & 0 \end{pmatrix} \cdot d \leq 0, \begin{pmatrix} 0 & 1 \end{pmatrix} \cdot d = 0 \right\} = \left\{ \begin{pmatrix} t \\ 0 \end{pmatrix} : t \geq 0 \right\} = T(X, \bar{x}).$$

The feasible set X can also be represented by

$$X = \{x \in \mathbb{R}^2 : x_2 - (x_1 + 1)^3 \leq 0, x_1 - 1 \leq 0, x_2 = 0\},$$

i. e.

$$g(x) = \begin{pmatrix} x_2 - (x_1+1)^3 \\ x_1 - 1 \end{pmatrix} \quad \text{and} \quad h(x) = x_2.$$

For $\bar{x}$, it holds that $\mathcal{A}(\bar{x}) = 1$, $\nabla g_1(\bar{x}) = \begin{pmatrix} -3(\bar{x}_1+1)^2 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ and $\nabla h(\bar{x}) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$.
Hence, it follows that

$$T_l(g,h,x) = \{d \in \mathbb{R}^2 : \begin{pmatrix} 0 & 1 \end{pmatrix} \cdot d \leq 0, \begin{pmatrix} 0 & 1 \end{pmatrix} \cdot d = 0\} = \left\{ \begin{pmatrix} t \\ 0 \end{pmatrix} : t \in \mathbb{R} \right\} \supsetneq T(X,x).$$

Next, we consider the good case where $T_l(g,h,x) = T(X,x)$.

Definition 15.11 The condition

$$T_l(g,h,x) = T(X,x)$$

is called **Abadie Constraint Qualification (ACQ)** for $x \in X$.

Then Theorem 15.6 can be simplified in the sense that the tangential cone $T(X, x)$ can be replaced by the simpler linearized tangential cone $T_l(g, h, x)$:

Theorem 15.12 (Necessary Optimality Condition) *Let $\bar{x}$ be a local minimum of (15.1) and let (ACQ) hold true for $\bar{x}$. Then it holds:*

a) $\bar{x} \in X$.

b) $\nabla f(\bar{x})^\top d \geq 0$ *for all* $d \in T_l(g, h, x)$.

Proof. It follows from Theorem 15.6. ∎

Next, we want to weaken the (ACQ) condition such that Theorem 15.12 still holds. For this purpose we define polar cones:

Definition 15.13 Let $C \subseteq \mathbb{R}^n$ with $C \neq \emptyset$ be a cone.

$$C^\circ := \{v \in \mathbb{R}^n : v^\top \cdot d \leq 0 \ \forall d \in C\}$$

is called **polar cone** of C.

Example 15.14

a) Let us illustrate polar cones in $\mathbb{R}^2$:
We note that $\langle y, x\rangle = \|x\| \cdot \|y\| \cdot \cos(\theta)$, where θ is the angle between x and y. Since $\cos(\theta)$ is non-positive for $\theta \in [\frac{\pi}{2}, \frac{3\pi}{2}]$, the angle between vectors in the cone and in the polar cone must lie between $\frac{\pi}{2}$ and $\frac{3\pi}{2}$.

b) Every linear subspace $L \subseteq \mathbb{R}^n$ is a cone. We consider its orthogonal complement

$$L^\perp := \{y \in \mathbb{R}^n : \langle y, x\rangle = 0 \ \forall x \in L\}.$$

If $y \in L^\perp$, then $\langle y, x\rangle = 0$ for all $x \in L$. Thus, it follows that $y \in L^\circ$, i. e. $L^\perp \subseteq L^\circ$.
If $y \in L^\circ$, then

$$\langle y, x\rangle \leq 0 \quad \forall x \in L. \tag{15.2}$$

Since L is a subspace, it holds for any $x \in L$ that $-x \in L$.
With (15.2), it follows that

$$\langle y, -x\rangle \leq 0 \quad \forall x \in L,$$

which is equivalent to

$$\langle y, x\rangle \geq 0 \quad \forall x \in L. \tag{15.3}$$

Thus, we can conclude from (15.2) and (15.3) that

$$\langle y, x\rangle = 0 \quad \forall x \in L,$$

i. e. $L^\circ \subseteq L^\perp$ and thus $L^\circ = L^\perp$.

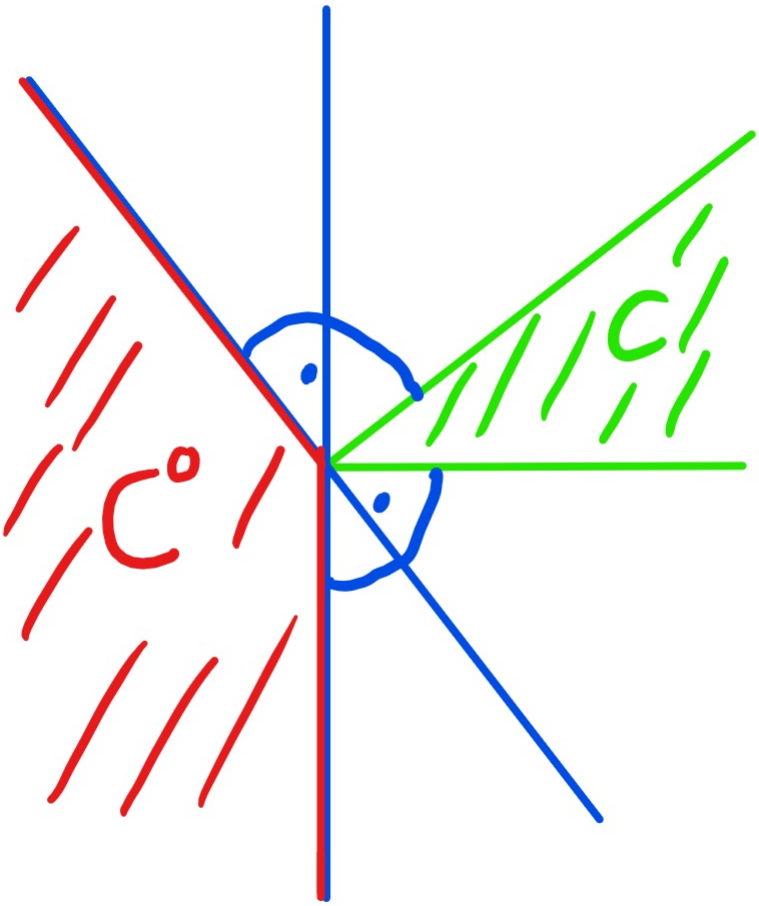


Figure 15.4.: Illustration of the polar cone

c) Let $b \in \mathbb{R}^n$. Then the half-line

$$h^+ := \{x \in \mathbb{R}^n : x = \lambda \cdot b, \lambda \geq 0\}$$

is a cone. We want to compute its polar cone. For this purpose, we can conclude that

$$\begin{aligned} y \in (h^+)^\circ \Leftrightarrow & \quad \langle y, \lambda \cdot b \rangle \leq 0 \quad \forall \lambda \geq 0 \\ \Leftrightarrow & \quad \lambda \cdot \langle y, b \rangle \leq 0 \quad \forall \lambda \geq 0 \\ \Leftrightarrow & \quad \langle y, b \rangle \leq 0 \\ \Leftrightarrow & y \in \{x \in \mathbb{R}^n : \langle x, b \rangle \leq 0\} \end{aligned}$$

This means that the polar cone of the half-line h^+ is a half-space.

By using the definition of a polar cone, we can rewrite the necessary optimality conditions

$$\nabla f(\bar{x})^\top \cdot d \geq 0 \quad \forall d \in T(X, \bar{x})$$

from Theorem 15.6 b) and

$$\nabla f(\bar{x})^\top \cdot d \geq 0 \quad \forall d \in T_l(g, h, \bar{x})$$

from Theorem 15.12 b) by

$$-\nabla f(\bar{x}) \in T(X, \bar{x})^\circ$$

and

$$-\nabla f(\bar{x}) \in T_l(g, h, \bar{x})^\circ,$$

respectively.
Thus, Theorem 15.12 follows from Theorem 15.6, if

$$T_l(g, h, x)^\circ = T(X, x)^\circ$$

holds. This leads to the following condition:

Definition 15.15 The condition

$$T_l(g, h, \bar{x})^\circ = T(X, \bar{x})^\circ$$

is called **Guignard Constraint Qualification (GCQ)** for $\bar{x} \in X$.

Our latter observations are summarized in the following two statements:

Lemma 15.16 *If (ACQ) holds for $x \in X$, then (GCQ) also holds for x.*

Theorem 15.17 (Necessary Optimality Condition) *Let $\bar{x}$ be a local minimum of (15.1). Let the (GCQ) condition be fulfilled for $\bar{x}$. Then it holds:*

a) $\bar{x} \in X$

b) $\nabla f(\bar{x})^\top \cdot d \geq 0 \quad \forall d \in T_l(g, h, \bar{x})$.

Later, we will derive sufficient conditions for (ACG)/(GCQ).

Definition 15.18 Let $x \in X$. A condition which implies (GCQ) is called **constraint qualification** for x.

For the next optimality condition, we need Farkas' Lemma from linear programming:

Theorem 15.19 (Farkas' Lemma) *Let $A \in \mathbb{R}^{n \times m}$, $B \in \mathbb{R}^{n \times p}$ and $c \in \mathbb{R}^n$.*
Then, the following two statements are equivalent:

a) For all $d \in \mathbb{R}^n$ with $A^\top \cdot d \leq 0$ and $B^\top \cdot d = 0$, it holds that $c^\top d \leq 0$.

b) There exists some $u \in \mathbb{R}^m$, $u \geq 0$ and $v \in \mathbb{R}^p$ with $c = Au + Bv$.

Proof. See the lecture 'Linear Optimization'. ∎

Now, we can state the next necessary optimality condition:

Theorem 15.20 (First-Order Necessary Conditions, KKT Conditions) *Let $\bar{x} \in \mathbb{R}^n$ be a local minimum of (15.1) for which a constraint qualification holds.*
Then, the Karush-Kuhn-Tucker conditions (KKT conditions) hold:
There exist Lagrangian multipliers $\bar{\lambda} \in \mathbb{R}^m$ and $\bar{\mu} \in \mathbb{R}^p$ such that:

a) $\nabla f(\bar{x}) + \nabla g(\bar{x}) \cdot \bar{\lambda} + \nabla h(\bar{x}) \cdot \bar{\mu} = 0$ *(multiplicator rule)*

b) $h(\bar{x}) = 0$,

c) $\bar{\lambda} \geq 0$, $g(\bar{x}) \leq 0$, $\bar{\lambda}^\top g(\bar{x}) = 0$ *(complementarity condition)*

Proof. Let $\bar{x}$ be a local minimum of (15.1) satisfying a constraint qualification.
Then, it follows that $h(\bar{x}) = 0$ and $g(\bar{x}) \leq 0$ and, by Theorem 15.12, it is $-\nabla f(\bar{x})^\top \cdot d \leq 0$ for all $d \in \mathbb{R}^n$ with $\nabla g_i(\bar{x})^\top \cdot d \leq 0$, $i \in \mathcal{A}(x)$, and $\nabla h(\bar{x})^\top \cdot d = 0$.
By applying Farkas' Lemma for $c = -\nabla f(\bar{x})$, $A = \nabla g_{\mathcal{A}(\bar{x})}(\bar{x})$ and $B = \nabla h(\bar{x})$, it follows that there exist $u \geq 0$ and $v \in \mathbb{R}^p$ with

$$c = Au + Bv.$$

Choosing $\bar{\lambda} \in \mathbb{R}^m$ with $\bar{\lambda}_{\mathcal{A}(\bar{x})} = u$, $\bar{\lambda}_{\mathcal{I}(\bar{x})} = 0$ and $\bar{\mu} = v$ shows the multiplicator rule.
Due to the choice of $\bar{\lambda}$, the complementarity condition is also satisfied. ∎

Definition 15.21 If $(\bar{x}, \bar{\lambda}, \bar{\mu}) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$ satisfies the KKT conditions, then $\bar{x}$ is called a **KKT point** and $(\bar{x}, \bar{\lambda}, \bar{\mu})$ is called a **KKT triple**.

Remark 15.22 In Theorem 15.20, part c) can be substituted by

c') $\bar{\lambda}_i \geq 0$, $g_i(\bar{x}) \leq 0$, $\bar{\lambda}_i g_i(\bar{x}) = 0$, $i \in U$

c") $g(\bar{x}) \leq 0$, $\bar{\lambda}_i \geq 0$ for $i \in \mathcal{A}(\bar{x})$, $\bar{\lambda}_i = 0$ for $i \in \mathcal{I}(\bar{x})$.

The complementarity condition ensures that $\bar{\lambda}_i = 0$ or $g_i(\bar{x}) = 0$ for every $i = 1, \ldots, m$.

Definition 15.23 Let $(\bar{x}, \bar{\lambda}, \bar{\mu})$ be a KKT triple.

a) **Strict complementarity** holds if $\bar{\lambda}_i > 0$ for all $i \in \mathcal{A}(\bar{x})$.

b) If there exists some $i \in U$ with $\bar{\lambda}_i = g_i(\bar{x}) = 0$, then **strict complementarity is violated**.

The multiplicator rule in the KKT conditions can be written in compact form by using the Lagrangian function:

Definition 15.24 The function $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ with

$$L(x, \lambda, \mu) = f(x) + \lambda^\top g(x) + \mu^\top h(x)$$

is called the **Lagrangian function**.

The multiplicator rule (Theorem 15.20) a)) is equivalent to

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

Additionally, the multiplicator rule together with the complementarity condition gives the interpretation that the negative gradient of the objective function f lies in the cone of active constraints.

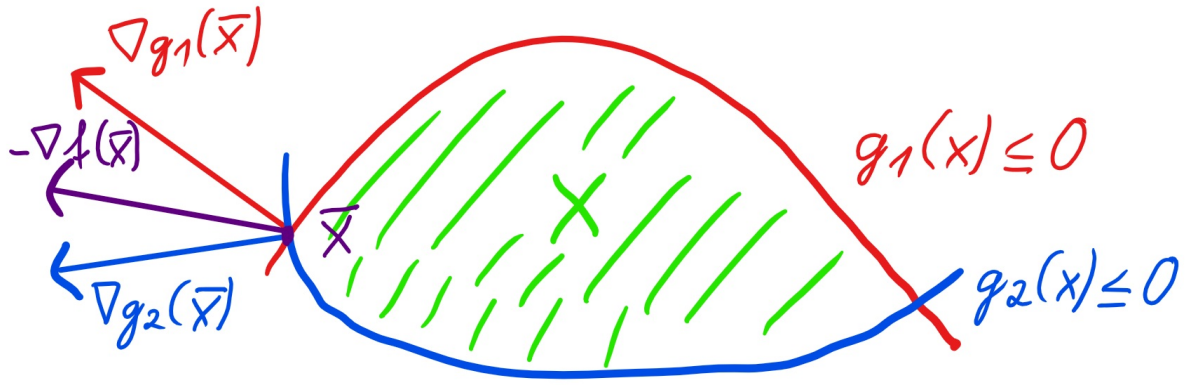


Figure 15.5.: Illustration of the restrictions on $\nabla f(x)$ in the KKT conditions

Constraint Qualifications

Next, we want to derive some constraint qualifications.

Definition 15.25 Let $X \subseteq \mathbb{R}^n$ be convex. The function $f : X \to \mathbb{R}$ is called **concave** if

$$f((1-\lambda)x + \lambda y) \geq (1-\lambda)f(x) + \lambda f(y) \quad \forall x, y \in X, \forall \lambda \in [0,1].$$

This is equivalent to the fact that $-f$ is convex.

We recall that a function $f : \mathbb{R}^n \to \mathbb{R}^p$ is called affine if $f(x) = Ax + b$ with $A \in \mathbb{R}^{p \times n}$ and $b \in \mathbb{R}^p$.

Theorem 15.26 *The following condition is a constraint qualification for $x \in X$:*

$$g_i \text{ concave}, i \in \mathcal{A}(x), \quad h \text{ affine} \tag{15.4}$$

Proof. Let $x \in X$ and (15.4) be satisfied.
We need to show that $T_l(g, h, x) \subseteq T(X, x)$.
Let $d \in T_l(g, h, x)$.
For $\alpha > 0$, we define $(\alpha_k)_k$ by $\alpha_k := \frac{\alpha}{k}$.
If α is sufficiently small, then

$$g_i(x + \alpha_k d) \leq 0 \quad \forall i \in \mathcal{I}(x), k \geq 1.$$

Since g_i is concave for $i \in \mathcal{A}(x)$, it follows that

$$g_i(x + \alpha_k d) \overset{\text{Thm. 6.3}a)}{\leq} g_i(x) + \alpha_k \nabla g_i(x)^\top d = \alpha_k \nabla g_i(x)^\top d \overset{\text{Def. 15.7}}{\leq} 0$$

Since h is affine, it follows that its Taylor polynomial of degree 1 is exact.
Hence, we can conclude that

$$h(x + \alpha_k d) = h(x) + \alpha_k \nabla h(x)^\top d = \alpha_k \nabla h(x)^\top d \overset{\text{Def. 15.7}}{=} 0.$$

Thus, $x^k = x + \alpha_k d \in X$ for all $k \geq 1$.
By choosing $\eta_k = \frac{1}{\alpha_k}$, it follows that $d \in T(X, x)$. ∎

Next, we consider another constraint qualification:

Definition 15.27 (MFCQ) We say that $x \in X$ satisfies the **Mangasarian-Fromovitz Constraint Qualification (MFCQ)** if

a) $\nabla h(x)$ has full column rank

b) there exists some $d \in \mathbb{R}^n$ with

$$\nabla g_i(x)^\top d < 0, i \in \mathcal{A}(x), \quad \nabla h(x)^\top d = 0.$$

If $m = 0$ or $\mathcal{A}(x) = \emptyset$, then b) does not apply.
If $p = 0$, then a) and $\nabla h(x)^\top d = 0$ in b) does not apply.

The condition (MFCQ) can be generalized in the following way:

Definition 15.28 (MFCQ') The (MFCQ') is defined by Definition 15.27, where part a) is replaced by

a') $\nabla h(x)$ has full column rank or h is affine.

Theorem 15.29 *If $x \in X$ satisfies (MFCQ) or its generalization (MFCQ'), then (ACQ) is also satisfied, i. e. (MFCQ) and (MFCQ') are both constraint qualifications.*

Next, we define a constraint qualification equivalent to (MFCQ):

Definition 15.30 $x \in X$ satisfies the **Positive Linear Independence Constraint Qualification (PLICQ)** if

a) $\nabla h(x)$ has full column rank

b) there are no vectors $u \in \mathbb{R}^m$, $v \in \mathbb{R}^p$ with

$$\nabla g(x) \cdot u + \nabla h(x) \cdot v = 0, \quad u_{\mathcal{A}(x)} \geq 0, \quad u_{\mathcal{A}(x)} \neq 0, \quad u_{\mathcal{I}(x)} = 0.$$

If $m = 0$ or $\mathcal{A}(x) = \emptyset$, then b) does not apply. If $p = 0$, then a) and the term $\nabla h(x)v$ in b) does not apply.

One can show that

$$x \in X \text{ satisfies (MFCQ)} \Leftrightarrow x \in X \text{ satisfies (PLICQ)}.$$

Definition 15.31 $x \in X$ is called **regular** if the columns of the matrix $(\nabla g_{\mathcal{A}(x)}, \nabla h(x))$ are linearly independent.

One can show that $x \in X$ being regular is a constraint qualification.

KKT conditions for convex problems

Next, we consider optimality conditions for convex problems.
We recall that a constrained nonlinear program is called convex if f, g_i, $i \in U$ are convex and h is affine. In this case, the feasible set is convex, because for all $x, y \in X$, $t \in [0,1]$, we get:

$$g_i(tx + (1-t)y) \leq tg_i(x) + (1-t)g_i(y) \leq 0, \quad i \in U$$
$$h(tx + (1-t)y) = th(x) + (1-t)h(y) = 0$$

Next, we will show that for convex nonlinear programs, the KKT conditions are necessary and sufficient:

Theorem 15.32 *Let the problem (15.1) be convex.*
Then every local minimum $\bar{x}$ of it is also a global minimum.
If the global minimum $\bar{x} \in X$ satisfies a constraint qualification, then $\bar{x}$ is a KKT point.
If $\bar{x} \in X$ is a KKT point, then $\bar{x}$ is an optimal solution of (15.1).

Proof. Let $\bar{x}$ be a local minimum of (15.1). Let $x \in X$.
For $d := x - \bar{x}$ and $t \in [0,1]$, it is $\bar{x} + td \in X$ and, for small values of t, it holds that

$$0 \leq f(\bar{x} + td) - f(\bar{x}) \leq (1-t)f(\bar{x}) + tf(x) - f(\bar{x}) = t(f(x) - f(\bar{x})).$$

Hence, it follows that $f(\bar{x}) \leq f(x)$ and that $\bar{x}$ is a global minimum.
From Theorem 15.20, it follows that $\bar{x}$ is a KKT point.
Let $\bar{x}$ be a KKT point, $x \in X$ and $d = x - \bar{x}$.
For $i \in U$, it holds that

$$\bar{\lambda}_i \nabla g_i(\bar{x})^\top \cdot d \overset{\text{convexity}}{\leq} \bar{\lambda}_i (g_i(x) - g_i(\bar{x})) \overset{\bar{\lambda}_i g_i(\bar{x}) = 0}{=} \underbrace{\bar{\lambda}_i}_{\geq 0} \underbrace{g_i(x)}_{\leq 0} \leq 0. \tag{15.5}$$

Moreover, it holds that

$$\nabla h(\bar{x})^\top \cdot d \overset{\text{Taylor}}{=} h(x) - h(\bar{x}) = 0. \tag{15.6}$$

Hence, it follows that

$$\begin{aligned} f(x) - f(\bar{x}) &\overset{f \text{ convex}}{\geq} \nabla f(\bar{x})^\top d \\ &\overset{\text{multiplicator rule}}{=} -\bar{\lambda}^\top \nabla g(\bar{x})^\top d - \bar{\mu}^\top \nabla h(\bar{x})^\top d \\ &\overset{(15.6)}{=} -\bar{\lambda}^\top \nabla g(\bar{x})^\top d \\ &\overset{(15.5)}{\geq} 0. \end{aligned}$$

Thus, $\bar{x}$ is a global minimum. ∎

Second-Order Optimality Condition

Next, we want to derive a second-order optimality condition for $(\bar{x}, \bar{\lambda}, \bar{\mu})$ such that $\bar{x}$ is a local minimum. For this purpose, we need to define another cone:

Definition 15.33 Let $x \in X$ and $\lambda \in [0, \infty)$. We define the cone

$$T_+(g,h,x,\lambda) := \left\{ d \in \mathbb{R}^n : \nabla g_i(x)^\top d \begin{cases} = 0, & \text{if } i \in \mathcal{A}(x) \text{ and } \lambda_i > 0 \\ \leq 0, & \text{if } i \in \mathcal{A}(x) \text{ and } \lambda_i = 0 \end{cases}, \nabla h(x)^\top d = 0 \right\}$$

Remark 15.34 It holds that $T_a(g,h,x) \subseteq T_+(g,h,x,\lambda) \subseteq T_l(g,h,x)$, where

$$T_a(g,h,x) = \{d \in \mathbb{R}^n : \nabla g_i(x)^\top d = 0, i \in \mathcal{A}(x), \nabla h(x)^\top d = 0\}$$

is the tangential space of active constraints.

Theorem 15.35 (Second-Order Sufficient Optimality Condition) *If $\bar{x} \in X$ is a KKT point with multipliers $\bar{\lambda} \in \mathbb{R}^m$ and $\bar{\mu} \in \mathbb{R}^p$ and if*

$$d^\top \nabla^2_{xx} L(\bar{x}, \bar{\lambda}, \bar{\mu}) d > 0 \quad \forall d \in T_+(g,h,\bar{x},\bar{\lambda}) \setminus \{0\}, \tag{15.7}$$

then $\bar{x}$ is an isolated local minimum of (15.1).

Proof. Assume $\bar{x}$ is not an isolated local minimum.
Then there exists some $(x^k)_k$ with $x^k \in X$, $x^k \neq \bar{x}$, $x^k \to \bar{x}$ and $f(x^k) \leq f(\bar{x})$.
Let $d^k = x^k - x$. Then a subsequence of $(y^k)_k$ with $y^k = \frac{d^k}{\|d^k\|}$ converges. Without loss of generality, let $(y^k)_k$ be this subsequence. Let y be its limit, i. e. $y^k \to y$.
It is

$$\frac{f(x^k) - f(\bar{x})}{\|d^k\|} = \nabla f(\bar{x})^\top y^k + \frac{o(\|d^k\|)}{\|d^k\|} \overset{k\to\infty}{\longrightarrow} \nabla f(\bar{x})^\top y$$

and

$$\frac{g_i(x^k) - g_i(\bar{x})}{\|d^k\|} \overset{k\to\infty}{\longrightarrow} \nabla g_i(\bar{x})^\top y, \quad \frac{h_i(x^k) - h_i(\bar{x})}{\|d^k\|} \overset{k\to\infty}{\longrightarrow} \nabla h_i(\bar{x})^\top y$$

Thus, it holds that

$$\nabla f(\bar{x})^\top y \leq 0, \quad \nabla g_i(\bar{x})^\top y \leq 0, i \in \mathcal{A}(x), \quad \nabla h_i(\bar{x})^\top y = 0.$$

From the KKT conditions, it follows that

$$0 = \nabla_x L(\bar{x}, \bar{\lambda}, \bar{\mu})^\top y = \underbrace{\nabla f(\bar{x})^\top y}_{\leq 0} + \sum_{i \in \mathcal{A}(x)} \bar{\lambda}_i \underbrace{\nabla g_i(\bar{x})^\top y}_{\leq 0} + \sum_{i=1}^{p} \bar{\mu}_i \underbrace{\nabla h_i(\bar{x})^\top y}_{=0}.$$

For all $i \in \mathcal{A}(\bar{x})$ with $\bar{\lambda}_i > 0$, it is $\nabla g_i(\bar{x})^\top y = 0$ since, otherwise, the right-hand side would be negative.

Hence, it holds that $y \in T_+(g, h, \bar{x}, \bar{\lambda})$. It follows that

$$\begin{aligned} L(x^k, \bar{\lambda}, \bar{\mu}) &= f(x^k) + \sum_{i \in \mathcal{A}(\bar{x})} \underbrace{\bar{\lambda}_i}_{\geq 0} \underbrace{g_i(x^k)}_{\leq 0} \\ &\leq f(x^k) \\ &\leq f(\bar{x}) \\ &\overset{\bar{\lambda}^\top g(\bar{x})=0;\, h(\bar{x})=0}{=} L(\bar{x}, \bar{\lambda}, \bar{\mu}) \end{aligned}$$

With $\nabla_x L(\bar{x}, \bar{\lambda}, \bar{\mu}) = 0$ and Taylor's Theorem, this leads to

$$0 \geq \frac{L(x^k, \bar{\lambda}, \bar{\mu}) - L(\bar{x}, \bar{\lambda}, \bar{\mu})}{\|d^k\|^2} = \frac{1}{2} {y^k}^\top \nabla^2_{xx} L(\bar{x}, \bar{\lambda}, \bar{\mu}) y^k + \frac{o(\|d\|^2)}{\|d\|^2} \overset{k \to \infty}{\longrightarrow} \frac{1}{2} y^\top \nabla^2_{xx} L(\bar{x}, \bar{\lambda}, \bar{\mu}) y,$$

which is a contradiction to (15.7).
Hence, the assumption was wrong and the claim follows. ∎

Example 15.36 Let $n = 2$, $m = 1$, $p = 0$, $f(x) = -x_1^2 + 2x_2$ and $g(x) = x_1^2 - x_2$. Then $\bar{x} = 0$ is the unique global minimum because for $x \in X \setminus \{0\}$, it holds that $x_2 > 0$, $x_1^2 \leq x_2$ and

$$f(x) = -x_1^2 + 2x_2 \geq -x_2 + 2x_2 = x_2 > 0 = f(0).$$

Additionally, we get that

$$\nabla f(\bar{x}) = \begin{pmatrix} 0 \\ 2 \end{pmatrix}, \quad \nabla^2 f(\bar{x}) = \begin{pmatrix} -2 & 0 \\ 0 & 0 \end{pmatrix}, \quad \nabla g(\bar{x}) = \begin{pmatrix} 0 \\ -1 \end{pmatrix} \text{ and } \nabla^2 g(\bar{x}) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}.$$

Thus, $\bar{x}$ is regular since $\nabla g(\bar{x}) \neq 0$.
$\bar{\lambda} = 2$ is the unique Lagrangian multiplier since

$$-\nabla f(\bar{x}) = \begin{pmatrix} 0 \\ -2 \end{pmatrix} = 2 \nabla g(\bar{x}).$$

Since $\bar{\lambda} > 0$, strict complementarity holds and

$$T_+(g, h, \bar{x}, \bar{\lambda}) = T_a(g, h, \bar{x}) = \left\{ \begin{pmatrix} \sigma \\ 0 \end{pmatrix} : \sigma \in \mathbb{R} \right\}.$$

The second-order sufficient optimality condition is satisfied because for all $d = \begin{pmatrix} \sigma \\ 0 \end{pmatrix}$ with $\sigma \neq 0$, it is

$$d^\top \nabla^2_{xx} L(\bar{x}, \bar{\lambda}) d = \sigma^2 \frac{d^2 L}{dx_1^2}(\bar{x}, \bar{\lambda}) = \sigma^2(-2 + 4) = 2\sigma^2 > 0.$$

We note that $\nabla^2 f(\bar{x})$ is negative definite on $T_+(g, h, \bar{x}, \bar{\lambda})$ since for $d = \begin{pmatrix} \sigma \\ 0 \end{pmatrix}$ with $\sigma \neq 0$, it holds that

$$d^\top \nabla^2 f(\bar{x}) d = \sigma^2 \frac{d^2 f}{dx_1^2}(\bar{x}) = -2\sigma^2 < 0.$$

We have just seen in this example that $\bar{x}$ is a global minimum that fulfills the second-order sufficient optimality condition, but $\nabla^2 f(\bar{x})$ is not positive (semi-)definite on $T_+(g, h, \bar{x}, \bar{\lambda})$. The reason for this behaviour is that $\nabla^2 f(\bar{x})$ only captures the curvature of f, but not of g. However, $\nabla^2_{xx} L(\bar{x}, \bar{\lambda}, \bar{\mu})$ takes the curvature of f and g into account.

Example 15.37 It holds that $T_a(g, h, \bar{x}) \subseteq T_+(g, h, \bar{x}, \bar{\lambda})$. But if $T_a(g, h, \bar{x}) \neq T_+(g, h, \bar{x}, \bar{\lambda})$, then positive definiteness of $\nabla^2_{xx} L(\bar{x}, \bar{\lambda}, \bar{\mu})$ on $T_a(g, h, \bar{x})$ is not enough to guarantee local optimality. This justifies the more complex notion of $T_+(g, h, \bar{x}, \bar{\lambda})$. An example for this scenario is given next:
Let $n = 2$, $m = 1$, $p = 0$, $f(x) = x_1^2 - x_2^2$ and $g(x) = x_1^2 - x_2$.
Then, for $\bar{x} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, it holds that $\nabla f(\bar{x}) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and $\nabla^2 f(\bar{x}) = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}$.
$\bar{\lambda} = 0$ is the only Lagrangian multiplier fulfilling the KKT conditions for $\bar{x}$.
From $\nabla g(\bar{x}) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$ it follows that

$$T_a(g, h, \bar{x}) = \{d \in \mathbb{R}^2 : \begin{pmatrix} 0 & -1 \end{pmatrix} \cdot d = 0\} = \left\{ \begin{pmatrix} \sigma \\ 0 \end{pmatrix} : \sigma \in \mathbb{R} \right\}.$$

For arbitrary $d = \begin{pmatrix} \sigma \\ 0 \end{pmatrix} \in T_a(g, h, \bar{x})$ with $\sigma \neq 0$, it is

$$d^\top \nabla^2_{xx} L(\bar{x}, \bar{\lambda}) d = 2\sigma^2 > 0.$$

But $\bar{x}$ is not a local minimum since for $x = \begin{pmatrix} 0 \\ \tau \end{pmatrix}$, $\tau > 0$, it is $x \in X$ and $f(x) = -\tau^2 < 0 = f(0)$.

16. Duality

In this chapter, we consider again the constrained nonlinear program

$$\begin{aligned} \min\ & f(x) \\ \text{s.t.}\ & g(x) \leq 0 \\ & h(x) = 0, \end{aligned} \tag{16.1}$$

and the Lagrangian function

$$L(x, \lambda, \mu) = f(x) + \lambda^\top g(x) + \mu^\top h(x).$$

In the duality theory for linear programs, dual solutions give bounds on primal objective function values and there exist algorithms inspired from duality, as e. g. the dual simplex method.
The question we want to answer in this chapter, is whether there exists a similar duality theory for nonlinear programs.
First, we note that

$$p(x) := \sup_{\substack{\lambda \in \mathbb{R}^m_+ \\ \mu \in \mathbb{R}^p}} L(x, \lambda, \mu) = \begin{cases} f(x) & \text{if } x \in X \\ \infty & \text{else} \end{cases}.$$

Thus our primal problem (16.1) is equivalent to

$$\min_{x \in \mathbb{R}^n} \sup_{\substack{\lambda \in \mathbb{R}^m_+ \\ \mu \in \mathbb{R}^p}} L(x, \lambda, \mu).$$

We obtain a dual problem to it by switching 'min' and 'sup':

Definition 16.1 The **dual problem** to (16.1) is given by the problem

$$\sup_{\substack{\lambda \in \mathbb{R}^m_+ \\ \mu \in \mathbb{R}^p}} \inf_{x \in \mathbb{R}^n} L(x, \lambda, \mu). \tag{16.2}$$

The function $d(\lambda, \mu) := \inf_{x \in \mathbb{R}^n} L(x, \lambda, \mu)$ is called **dual objective function**. The function $p(x) := \sup_{\substack{\lambda \in \mathbb{R}^m_+ \\ \mu \in \mathbb{R}^p}} L(x, \lambda, \mu)$ is called **primal objective function**.

Remark 16.2 Let $x \in \mathbb{R}$. Then $(\lambda, \mu) \to L(x, \lambda, \mu)$ is a linear function. The function $d(\lambda, \mu)$ is the infimum of linear functions and thus concave. The dual problem is a maximization problem over a convex set and thus equivalent to a convex problem.

The following theorem shows that the dual problem delivers lower bounds for the primal problem:

Theorem 16.3 (Weak Duality) *Let $\widetilde{x}$ be feasible for (16.1) and $(\widetilde{\lambda}, \widetilde{\mu})$ feasible for the dual problem (16.2). Then it holds that*

$$p(\widetilde{x}) = f(\widetilde{x}) \geq d(\widetilde{\lambda}, \widetilde{\mu}).$$

Proof. It holds that

$$d(\widetilde{\lambda}, \widetilde{\mu}) = \inf_{x \in \mathbb{R}^n} L(x, \lambda, \mu) \leq L(\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}) = f(\widetilde{x}) + \widetilde{\lambda}^\top g(\widetilde{x}) + \widetilde{\mu}^\top h(\widetilde{x}) \leq f(\widetilde{x}) = p(\widetilde{x}). \quad \blacksquare$$

The next question we want to answer is: Under which conditions does equality of the objective function values hold?
For this purpose, we define saddle points:

Definition 16.4 $(\bar{x}, \bar{\lambda}, \bar{\mu}) \in \mathbb{R}^n \times \mathbb{R}^m_+ \times \mathbb{R}^p$ is called **saddle point** of $L(x, \lambda, \mu)$ if

$$L(\bar{x}, \lambda, \mu) \leq L(\bar{x}, \bar{\lambda}, \bar{\mu}) \leq L(x, \bar{\lambda}, \bar{\mu}) \quad \forall x \in \mathbb{R}^n, \lambda \in \mathbb{R}^m_+, \mu \in \mathbb{R}^p.$$

Then we get the following strong duality result:

Theorem 16.5 (Strong Duality) *The following two statements are equivalent:*

a) $(\bar{x}, \bar{\lambda}, \bar{\mu})$ *is a saddle point of* $L(x, \lambda, \mu)$.

b) $\bar{x}$ *is a global minimum of (16.1),* $(\bar{\lambda}, \bar{\mu})$ *is a global maximum of (16.2) and* $f(\bar{x}) = d(\bar{\lambda}, \bar{\mu})$.

Proof. "$a) \Rightarrow b)$": For any $\widetilde{x} \in \mathbb{R}^n$, it holds that

$$\sup_{\substack{\lambda \in \mathbb{R}^m_+ \\ \mu \in \mathbb{R}^p}} \inf_{x \in \mathbb{R}^n} L(x, \lambda, \mu) \leq \sup_{\substack{\lambda \in \mathbb{R}^m_+ \\ \mu \in \mathbb{R}^p}} L(\widetilde{x}, \lambda, \mu).$$

Since this holds for any $\widetilde{x}$, it follows that

$$\sup_{\substack{\lambda \in \mathbb{R}^m_+ \\ \mu \in \mathbb{R}^p}} \inf_{x \in \mathbb{R}^n} L(x, \lambda, \mu) \leq \inf_{x \in \mathbb{R}^n} \sup_{\substack{\lambda \in \mathbb{R}^m_+ \\ \mu \in \mathbb{R}^p}} L(x, \lambda, \mu). \tag{16.3}$$

We get that

$$\begin{aligned}
L(\bar{x}, \bar{\lambda}, \bar{\mu}) &\overset{a)}{=} \inf_{x \in \mathbb{R}^n} L(x, \bar{\lambda}, \bar{\mu}) \\
&\leq \sup_{\substack{\lambda \in \mathbb{R}^m_+ \\ \mu \in \mathbb{R}^p}} \inf_{x \in \mathbb{R}^n} L(x, \lambda, \mu) \\
&\overset{(16.3)}{\leq} \inf_{x \in \mathbb{R}^n} \sup_{\substack{\lambda \in \mathbb{R}^m_+ \\ \mu \in \mathbb{R}^p}} L(x, \lambda, \mu) \\
&\leq \sup_{\substack{\lambda \in \mathbb{R}^m_+ \\ \mu \in \mathbb{R}^p}} L(\bar{x}, \lambda, \mu) \\
&\overset{a)}{=} L(\bar{x}, \bar{\lambda}, \bar{\mu}).
\end{aligned}$$

Thus, it follows that

$$L(\bar{x}, \bar{\lambda}, \bar{\mu}) = \inf_{x \in \mathbb{R}^n} L(x, \bar{\lambda}, \bar{\mu}) = d(\bar{\lambda}, \bar{\mu}) = \sup_{\substack{\lambda \in \mathbb{R}^m_+ \\ \mu \in \mathbb{R}^p}} L(\bar{x}, \lambda, \mu) = p(\bar{x}) < \infty.$$

Hence, it holds that $\bar{x}$ is feasible and that $d(\bar{\lambda}, \bar{\mu}) = p(\bar{x}) = f(\bar{x})$. The optimality of $\bar{x}$ for (16.1) and $(\bar{\lambda}, \bar{\mu})$ for (16.2) follows from the weak duality.
"$b) \Rightarrow a)$":
From b) and the definitions of p and d, it follows that

$$L(\bar{x}, \bar{\lambda}, \bar{\mu}) \leq f(\bar{x}) = p(\bar{x}) = \sup_{\substack{\lambda \in \mathbb{R}^m_+ \\ \mu \in \mathbb{R}^p}} L(\bar{x}, \lambda, \mu) = d(\bar{\lambda}, \bar{\mu}) = \inf_{x \in \mathbb{R}^n} L(x, \bar{\lambda}, \bar{\mu}) \leq L(\bar{x}, \bar{\lambda}, \bar{\mu}).$$

Thus, it holds that

$$\sup_{\substack{\lambda \in \mathbb{R}^m_+ \\ \mu \in \mathbb{R}^p}} L(\bar{x}, \lambda, \mu) = L(\bar{x}, \bar{\lambda}, \bar{\mu}) = \inf_{x \in \mathbb{R}^n} L(x, \bar{\lambda}, \bar{\mu}),$$

i. e. $(\bar{x}, \bar{\lambda}, \bar{\mu})$ is a saddle point of the Lagrangian function. ∎

For continuously differentiable, convex nonlinear programs (16.1), the dual problem can often be stated explicitely. Under these assumptions, the function

$$x \mapsto L(x, \lambda, \mu)$$

is convex and

$$\inf_{x \in \mathbb{R}^n} L(x, \bar{\lambda}, \bar{\mu}) = \min_{x \in \mathbb{R}^n} L(x, \bar{\lambda}, \bar{\mu}) \quad \text{for all } (\lambda, \mu).$$

These minimal points are exactly those with

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

Thus, the dual problem can be rewritten in this case to

$$\begin{aligned} \sup \; & L(x, \lambda, \mu) \\ \text{s.t.} \; & \lambda \geq 0 \\ & \nabla_x L(x, \lambda, \mu) = 0, \end{aligned}$$

This problem is called **Wolfe dual problem**.

17. Penalty Methods

In this chapter we consider so-called penalty methods for solving constrained nonlinear programs. The idea of these methods is to solve a constrained nonlinear program by a sequence of unconstrained nonlinear programs. These unconstrained nonliner programs use penalty terms in their objective functions. The penalty terms shall penalize the violation of feasibility, i. e. these unconstrained nonlinear programs are so-called Penalty subproblems

$$\min \lim_{x \in \mathbb{R}^n} f(x) + \alpha\pi(x),$$

where $\alpha > 0$ is a penalty parameter and $\pi : \mathbb{R}^n \to \mathbb{R}$ a penalty function with $\pi(x) = 0$ for $x \in X$ and $\pi(x) > 0$ for $x \in \mathbb{R}^n \setminus X$.
The bigger the penalty parameter α is, the better approximates the penalty subproblem the original constrained nonlinear program. But a larger α makes the penalty subproblem numerically more difficult.
Thus, one solves a sequence of penalty problems which correspond to a monotonically increasing sequence of penalty parameters. The optimal solution of the previous problem is then used as the starting point for the next problem.

Quadratic Penalty Methods

First, we consider quadratic penalty methods. They use **quadratic penalty functions**

$$\begin{aligned} P_\alpha(x) &= f(x) + \frac{\alpha}{2} \sum_{i=1}^{m} \max^2\{0, g_i(x)\} + \frac{\alpha}{2} \sum_{i=1}^{p} h_i(x)^2 \\ &= f(x) + \frac{\alpha}{2} \|(g(x))_+\|^2 + \frac{\alpha}{2} \|h(x)\|^2, \end{aligned}$$

where $\alpha > 0$ is the penalty parameter and the two penalty terms

$$p_u(t) = (t)_+^2 = \max^2\{0, t\} \quad \text{for } g_i(x) \leq 0$$

and

$$p_g(t) = t^2 \quad \text{for } h_i(x) = 0$$

are used.
The quadratic penalty function P_α is continuously differentiable, because p_u and p_g are both continuously differentiable. However, the drawback of the penalty terms p_u and p_g is that

$$p_u'(t) = 2(t)_+ \quad \text{and} \quad p_g'(0) = 0,$$

i .e the slope of both penalty terms is zero in X. More precisely, we get

$$\nabla P_\alpha(x) = \nabla f(x) + \alpha \sum_{i=1}^{m} (g_i(x))_+ \nabla g_i(x) + \alpha \sum_{i=1}^{p} h_i(x) \nabla h_i(x).$$

Especially, this means that

$$P_\alpha(x) = f(x) \quad \text{and} \quad \nabla P_\alpha(x) = \nabla f(x) \quad \forall x \in X.$$

In the particular moment where the feasible set is left, the slope of the penalty terms is therefore zero, i. e. the penalty has only a delayed effect. Additionally, a point $x \in X$ is only a stationary point of P_α if $\nabla f(x) = 0$. But usually, this is not the case for a point $x \in X$, i. e. usually, minimzing the penalty function will lead to an infeasible point.
The Penalty Method has then the following form:

Algorithm 17.1 Penalty Method

1: Choose $\alpha > 0$
2: **for** $k = 0, 1, 2, \ldots$ **do**
3: Compute global optimal solution x^k of the penalty problem

$$\min_{x \in \mathbb{R}^n} P_{\alpha_k}(x)$$

(if $k > 0$, choose x^{k-1} as starting point)
4: **if** $x^k \in X$ **then**
5: Stop
6: Choose $\alpha_{k+1} > \alpha_k$

We get the following convergence result:

Theorem 17.2 *Let f, g, h be continuous, $X \neq \emptyset$. Let $(\alpha_k)_k \subset (0, \infty)$ be strictly monotonically increasing with $\alpha_k \to \infty$. Let Algorithm 17.1 generate the sequence $(x^k)_k$. Then it holds:*

a) $(P_{\alpha_k}(x^k))_k$ is monotonically increasing.

b) $(\|(g(x^k))_+\|^2 + \|h(x^k)\|^2)_k$ is monotonically decreasing.

c) $(f(x^k))_k$ is monotonically increasing

d) $\lim_{k\to\infty} (g(x^k))_+ = 0$ and $\lim_{k\to\infty} h(x^k) = 0$.

e) Every accumulation point of $(x^k)_k$ is a global optimal solution of (15.1).

Proof. We define $\pi(x) := \frac{1}{2}(\|(g(x))_+\|^2 + \|h(x)\|^2)$.

a) It follows that

$$P_{\alpha_k}(x^k) \leq P_{\alpha_k}(x^{k+1}) = f(x^{k+1}) + \alpha_k \pi(x^{k+1}) \leq f(x^{k+1}) + \alpha_{k+1}\pi(x^{k+1}) = P_{\alpha_{k+1}}(x^{k+1}).$$

b) Adding $P_{\alpha_k}(x^k) \leq P_{\alpha_k}(x^{k+1})$ and $P_{\alpha_{k+1}}(x^{k+1}) \leq P_{\alpha_{k+1}}(x^k)$ leads to

$$\alpha_k \pi(x^k) + \alpha_{k+1}\pi(x^{k+1}) \leq \alpha_k \pi(x^{k+1}) + \alpha_{k+1}\pi(x^k)$$

With $\alpha_k < \alpha_{k+1}$ it follows that $\pi(x^k) \geq \pi(x^{k+1})$.

c) It holds that

$$0 \leq P_{\alpha_k}(x^{k+1}) - P_{\alpha_k}(x^k) = f(x^{k+1}) - f(x^k) + \alpha_k \underbrace{(\pi(x^{k+1}) - \pi(x^k))}_{\leq 0 \text{ due to } b)} \leq f(x^{k+1}) - f(x^k).$$

d) We show that $\pi(x^k) \to 0$. Since $X \neq \emptyset$, there exists some $\hat{x} \in X$. It follows that

$$f(\hat{x}) = P_{\alpha_k}(\hat{x}) \geq P_{\alpha_k}(x^k) = f(x^k) + \alpha_k \pi(x^k) \overset{c)}{\geq} f(x^0) + \alpha_k \pi(x^k).$$

Since $\alpha_k \to \infty$, it is $\pi(x^k) \to 0$.

e) Let $\bar{x}$ be an accumulation point of $(x^k)_k$. Then, $\bar{x} \in X$ due to d) and since $(g)_+$ and h are continuous. Let $(x^k)_K$ with $x^k \to \bar{x}$, $k \in K$, $k \to \infty$.
Then it holds for all $x \in X$, $k \in K$ that

$$f(x^k) \leq P_{\alpha_k}(x^k) \leq P_{\alpha_k}(x) = f(x).$$

It follows that

$$f(\bar{x}) = \lim_{\substack{k \to \infty \\ k \in K}} f(x^k) \leq f(x) \quad \forall x \in X.$$ ∎

In Theorem 17.2, it is assumed that Algorithm 17.1 generates an infinite sequence $(x^k)_k$. If this is not the case, then one of the subproblems was not solvable or the stopping criterion $x^k \in X$ was fulfilled for some k. In the latter case, it holds that x^k is a global optimal solution of (15.1) since

$$f(x) = P_{\alpha_k}(x) \geq P_{\alpha_k}(x^k) = f(x^k) \quad \forall x \in X.$$

Next we want to consider relation of KKT points and the penalty method. For this purpose, let f, g, h be continuously differentiable. Let $\bar{x}$ be an accumulation point of $(x^k)_k$ with $\alpha_k \to \infty$. Theorem 17.2 e) states that $\bar{x}$ is a global minimum of (15.1). For every k, the point x^k is a stationary point of P_{α_k}, i. e.

$$\begin{aligned} 0 &= \nabla P_{\alpha_k}(x^k) \\ &= \nabla f(x^k) + \sum_{i=1}^{m} \alpha_k \max\{0, g_i(x^k)\} \nabla g_i(x^k) + \sum_{i=1}^{p} \alpha_k h_i(x^k) \nabla h_i(x^k) \\ &= \nabla f(x^k) + \nabla g(x^k) \lambda^k + \nabla h(x^k) \mu^k \end{aligned}$$

with

$$\lambda_i^k = \alpha_k \max\{0, g_i(x^k)\} \quad \text{and} \quad \mu_i^k = \alpha_k h_i(x^k). \tag{17.1}$$

Equivalently, this means that

$$\nabla_x L(x^k, \lambda^k, \mu^k) = 0. \tag{17.2}$$

If there exists a subsequence with $(x^k)_K \to \bar{x}$, $(\lambda^k)_K \to \bar{\lambda}$ and $(\mu^k)_K \to \bar{\mu}$, then $(\bar{x}, \bar{\lambda}, \bar{\mu})$ is a KKT triple:

Theorem 17.3 *Let f, g, h be continuously differentiable, $X \neq \emptyset$. Let $(\alpha_k)_k \subset (0, \infty)$ be strictly monotonically increasing with $\alpha_k \to \infty$. Suppose Algorithm 17.1 generates the sequence $(x^k)_k$. We define $(\lambda^k)_k$ and $(\mu^k)_k$ as in (17.1). Then it holds:*

a) If $(x^k, \lambda^k, \mu^k)_K$ is a convergent subsequence of $(x^k, \lambda^k, \mu^k)_k$ with limit $(\bar{x}, \bar{\lambda}, \bar{\mu})$, then $\bar{x}$ is $\bar{x}$ is a global optimal solution of (15.1) and $(\bar{x}, \bar{\lambda}, \bar{\mu})$ is a KKT triple of (15.1).

b) Let $\bar{x}$ be an accumulation point of $(x^k)_k$ and let $(x^k)_K$ be a subsequence converging to $\bar{x}$. Let $\bar{x}$ be regular. Then the sequence $(x^k, \lambda^k, \mu^k)_K$ converges to a KKT triple of (15.1) and $\bar{x}$ is a global optimal solution of (15.1).

Proof.

a) From Theorem 17.2 e), it follows that $\bar{x}$ is a global optimal solution of $\bar{x}$. Taking the limit $K \ni k \to \infty$ in (17.2) leads to

$$0 = \lim_{\substack{k \to \infty \\ k \in K}} \nabla_x L(x^k, \lambda^k, \mu^k) = \nabla_x L(\bar{x}, \bar{\lambda}, \bar{\mu}).$$

Thus, the multiplicator rule holds. We know from the optimality that $\bar{x} \in X$. The complementarity condition is left to show. Due to (17.1), it follows that $\lambda^k \geq 0$ and therefore $\lambda \geq 0$.
If $g_i(\bar{x}) < 0$, it follows that $g_i(x^k) < 0$ for $k \in K$ sufficiently large. Hence, it follows that $\lambda_i^k = \alpha_k \max\{0, g_i(x^k)\} = 0$ and therefore $\bar{\lambda}_i = 0$. Thus, complementarity holds.

b) As in a), it follows that $\bar{x}$ is a global optimal solution of (15.1). It is left to show that $(x^k, \lambda^k, \mu^k)_K$ converges (the rest follows from a)).
For $i \in \mathcal{I}(\bar{x})$, it holds that $g_i(\bar{x}) < 0$.
Hence, $g_i(x^k) < 0$ for $k \in K$ sufficiently large.
It follows that $\lambda_i^k = \alpha_k \max\{0, g_i(x^k)\} = 0$ for k sufficiently large and that

$$\bar{\lambda}_{\mathcal{I}(\bar{x})} := \lim_{K \ni k \to \infty} \lambda^k_{\mathcal{I}(\bar{x})} = 0.$$

Since $\bar{x}$ is regular, it holds that $A_\star := (\nabla g_{\mathcal{A}(\bar{x})}(\bar{x}), \nabla h(\bar{x}))$ has full column rank.
Hence, $A_\star^\top A_\star$ is invertible.
Due to continuity, it follows that $A_k^\top A_k$ is also invertible (for large k), where

$A_k = (\nabla g_{\mathcal{A}(\bar{x})}(x^k), \nabla h(x^k))$.
Hence, we get for sufficiently large k that

$$0 = A_k^\top \nabla_x L(x^k, \lambda^k, \mu^k) = A_k^\top \nabla f(x^k) + A_k^\top A_k \begin{pmatrix} \lambda^k_{\mathcal{A}(\bar{x})} \\ \mu^k \end{pmatrix}$$

and that

$$\begin{pmatrix} \lambda^k_{\mathcal{A}(\bar{x})} \\ \mu^k \end{pmatrix} = -(A_k^\top A_k)^{-1} A_k^\top \nabla f(x^k) \overset{k\to\infty, k\in K}{\longrightarrow} -(A_\star^\top A_\star)^{-1} A_\star^\top \nabla f(\bar{x}).$$

Thus, (x^k, λ^k, μ^k) converges and the result follows with a). ∎

Remark 17.4 We recall that $\alpha_k \to \infty$. This can lead to a bad condition of the penalty problems. To see this, we consider equality-constrained problems with f, h two-times continuously differentiable. Then it holds that

$$\nabla^2 P_\alpha(x) = \nabla^2 f(x) + \alpha \nabla h(x) \nabla h(x)^\top + \alpha \sum_{i=1}^{p} h_i(x) \nabla^2 h_i(x).$$

Moreover, let h be affine, i. e. $h(x) = A^\top x + b$ with $A \in \mathbb{R}^{n\times p} \setminus \{0\}$ and $b \in \mathbb{R}^p$.
Then we get that

$$\nabla^2 P_\alpha(x) = \nabla^2 f(x) + \alpha A A^\top.$$

Let v be arbitrary with $A^\top v \neq 0$. Then it holds that

$$v^\top \nabla^2 P_\alpha(x) v = v^\top \nabla^2 f(x) v + \alpha \|A^\top v\|^2 = \mathcal{O}(\alpha) \quad (\alpha \to \infty)$$

For $w \neq 0$ with $A^\top w = 0$, it holds that

$$w^\top \nabla^2 P_\alpha(x) w = w^\top \nabla^2 f(x) w = \mathcal{O}(1).$$

Hence, the condition of $\nabla^2 P_\alpha(x)$ for $\alpha \to \infty$ is in $\mathcal{O}(\alpha)$. This causes problems in gradient-based methods and leads to small regions of fast local convergence in Newton-based methods.

Exact Penalty Methods

Exact penalty methods use penalty terms with the following property:

Definition 17.5 Let $\bar{x} \in \mathbb{R}^n$ be a local optimal solution of (15.1). The penalty function $P : \mathbb{R}^n \to \mathbb{R}$ is called **exact** in $\bar{x}$ if $\bar{x}$ is a local minimum of P.

Example 17.6 The **l1-penalty function**

$$P_\alpha^1(x) = f(x) + \alpha \sum_{i=1}^{m} (g_i(x))_+ + \alpha \sum_{i=1}^{p} |h_i(x)| = f(x) + \alpha(\|(g_i(x))_+\|_1 + \|h(x)\|_1)$$

is exact. Its drawback is that it is not differentiable because $(\cdot)_+$ and $|\cdot|$ are not differentiable.

The next theorem shows the exactness of P_α^1 for convex optimization problems:

Theorem 17.7 *Let $(\bar{x}, \bar{\lambda}, \bar{\mu})$ be a KKT triple of (15.1) with convex, continuously differentiable functions $f, g_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \dots, m$ and an affine function $h : \mathbb{R}^n \to \mathbb{R}^p$. Then, $\bar{x}$ is a global optimal solution of (15.1) and $\bar{x}$ is a global minimum of P_α^1 on $\mathbb{R}^n$ for all*

$$\alpha \geq \max\{\bar{\lambda}_1, \dots, \bar{\lambda}_m, |\bar{\mu}_1|, \dots, |\bar{\mu}_p|\}.$$

Proof. Global optimality follows from the fact that $\bar{x}$ is a KKT point, f, g are convex, h is affine and Theorem 15.32.
Since $\bar{\lambda} \geq 0$, it follows that $L(\cdot, \bar{\lambda}, \bar{\mu})$ is convex.
With $\nabla_x L(\bar{x}, \bar{\lambda}, \bar{\mu}) = 0$ and Theorem 6.3, we can conclude that

$$L(x, \bar{\lambda}, \bar{\mu}) \geq L(\bar{x}, \bar{\lambda}, \bar{\mu}) + \nabla_x L(\bar{x}, \bar{\lambda}, \bar{\mu})(x - \bar{x}) = L(\bar{x}, \bar{\lambda}, \bar{\mu}) \ \forall x \in \mathbb{R}^n \qquad \blacksquare$$

Hence, $\bar{x}$ is a global minimum of $L(\cdot, \bar{\lambda}, \bar{\mu})$. Together with $\bar{x} \in X$ it follows for all $x \in \mathbb{R}^n$ that

$$\begin{aligned}
P_\alpha^1(\bar{x}) &= f(\bar{x}) + \alpha\|(g(\bar{x}))_+\|_1 + \alpha\|h(\bar{x})\|_1 \\
&= f(\bar{x}) \\
&= f(\bar{x}) + \bar{\lambda}^\top g(\bar{x}) + \bar{\mu}^\top h(\bar{x}) \\
&= L(\bar{x}, \bar{\lambda}, \bar{\mu}) \\
&\leq L(x, \bar{\lambda}, \bar{\mu}) \\
&= f(x) + \sum_{i=1}^{m} \bar{\lambda}_i g_i(x) + \sum_{i=1}^{p} \bar{\mu}_i h_i(x) \\
&\leq f(x) + \sum_{i=1}^{m} \bar{\lambda}_i (g_i(x))_+ + \sum_{i=1}^{p} |\mu_i| \cdot |h_i(x)| \\
&\leq f(x) + \sum_{i=1}^{m} \alpha (g_i(x))_+ + \sum_{i=1}^{p} \alpha |h_i(x)| \\
&= P_\alpha^1(x)
\end{aligned}$$

18. Sequential Quadratic Programming

Sequential Quadratic Programming (SQP) is one of the most efficient optimization methods.

Lagrange-Newton Method for Equality Constraints

First, we explain the concept of (SQP) for problems of the form

$$\begin{aligned} &\min\ f(x) \\ &\text{s.t.}\ \ h(x) = 0. \end{aligned} \tag{18.1}$$

Let $\bar{x}$ be a local minimum of (18.1) which satisfies a constraint qualification. Then it follows from Theorem 15.20 that the KKT conditions hold, which have in this case the following form:

$$\begin{aligned} \exists \bar{\mu} \in \mathbb{R}^p : \nabla_x L(\bar{x}, \bar{\mu}) &= 0 \\ h(\bar{x}) &= 0. \end{aligned} \tag{18.2}$$

This system of equations consists of $n+p$ variables and $n+p$ equations. Hence, the idea is to apply Newton's method to

$$F(x, \mu) := \begin{pmatrix} \nabla_x L(x, \mu) \\ h(x) \end{pmatrix} = 0 \tag{18.3}$$

in order to get the KKT pair $(\bar{x}, \bar{\mu})$.
For being able to apply Newton's method, we assume that f and h are two-times continuously differentiable. Then, F is continuously differentiable and

$$F'(x, \mu) = \begin{pmatrix} \nabla^2_{xx} L(x, \mu) & \nabla^2_{x\mu} L(x, \mu) \\ \nabla h(x)^\top & 0 \end{pmatrix} = \begin{pmatrix} \nabla^2_{xx} L(x, \mu) & \nabla h(x) \\ \nabla h(x)^\top & 0 \end{pmatrix}.$$

Let (x^k, μ^k) be the current point in Newton's method. Then the Newton step d^k for (18.3) is given by

$$F'(x^k, \mu^k) d^k = -F(x^k, \mu^k).$$

Explicitely, this is the so-called **Lagrange-Newton equation (LNE)**

$$\begin{pmatrix} \nabla^2_{xx} L(x, \mu) & \nabla h(x) \\ \nabla h(x)^\top & 0 \end{pmatrix} \begin{pmatrix} d^k_x \\ d^k_\mu \end{pmatrix} = - \begin{pmatrix} \nabla_x L(x^k, \mu^k) \\ h(x^k) \end{pmatrix} \tag{18.4}$$

with $d^k = \begin{pmatrix} d_x^k \\ d_\mu^k \end{pmatrix} \in \mathbb{R}^n \times \mathbb{R}^p$.
This leads to the following algorithm:

Algorithm 18.1 Lagrange-Newton Method

1: Choose $x^0 \in \mathbb{R}^n$ and $\mu^0 \in \mathbb{R}^p$
2: **for** $k = 0, 1, 2, \ldots$ **do**
3: **if** $h(x^k) = 0$ and $\nabla_x L(x^k, \mu^k) = 0$ **then**
4: Stop
5: Compute $d^k = \begin{pmatrix} d_x^k \\ d_\mu^k \end{pmatrix}$ by solving the Lagrange-Newton equation (18.4)
6: $x^{k+1} := x^k + d_x^k$; $\mu^{k+1} := \mu^k + d_\mu^k$

To show the local convergence of this method, we want to apply the local convergence Theorem 10.5 for Newton's method. This theorem requires the assumption that

$$F'(x, \mu) = \begin{pmatrix} \nabla_{xx}^2 L(x, \mu) & \nabla h(x) \\ \nabla h(x)^\top & 0 \end{pmatrix}$$

is invertible in the KKT pair $(\bar{x}, \bar{\mu})$. For this purpose, we can use the following characterization:

Lemma 18.2 *Let f and h be two-times continuously differentiable. Let $x \in \mathbb{R}^n$ and $\mu \in \mathbb{R}^p$. If rank$(\nabla h(x)) = p$ and $s^\top \nabla_{xx}^2 L(x, \mu) s > 0$ for all $s \in \mathbb{R}^n \setminus \{0\}$ with $\nabla h(x)^\top s = 0$, then the matrix*

$$\begin{pmatrix} \nabla_{xx}^2 L(x, \mu) & \nabla h(x) \\ \nabla h(x)^\top & 0 \end{pmatrix}$$

is invertible.

Proof. We show: If $F'(x, y) \begin{pmatrix} v \\ w \end{pmatrix} = 0$, then $\begin{pmatrix} v \\ w \end{pmatrix} = 0$.
(Then, F' is injective and therefore as a quadratic matrix also invertible.)
From the second block row, it follows that

$$\nabla h(x)^\top v = 0. \tag{18.5}$$

Multiplying the first block row with $v^\top$ from the left leads to

$$0 = \underbrace{v^\top \nabla_{xx}^2 L(x, \mu) v}_{>0, \text{ if } v \neq 0} + \underbrace{v^\top \nabla h(x)}_{=0 \text{ by } (18.5)} w$$

Hence it holds that $v = 0$ and it follows from the first block row that

$$\nabla h(x) w = 0.$$

Since the columns of $\nabla h(x)$ are linearly independent, it follows that $w = 0$. ∎

Next we obtain a local convergence result for Algorithm 18.1 by applying Theorem 10.5:

Theorem 18.3 *Let f and h be two-times continuously differentiable. Let $(\bar{x}, \bar{\mu})$ be a KKT pair with*

1. *rank$(\nabla h(\bar{x})) = p$*
2. *$s^\top \nabla^2_{xx} L(\bar{x}, \bar{\mu}) s > 0 \ \forall s \in \mathbb{R}^n \setminus \{0\}$ with $\nabla h(\bar{x})^\top s = 0$.*

Then there is a $\delta > 0$ such that for all $(x^0, \mu^0) \in B_\delta(\bar{x}, \bar{\mu})$, Algorithm 18.1 either terminates with $(x^k, \mu^k) = (\bar{x}, \bar{\mu})$ or generates a sequence (x^k, μ^k) which converges q-superlinearly to $(\bar{x}, \bar{\mu})$:

$$\|(x^{k+1} - \bar{x}, \mu^{k+1} - \bar{\mu})\| = o(\|(x^k - \bar{x}, \mu^k - \bar{\mu})\|) \quad (k \to \infty)$$

Moreover, if $\nabla^2 f$ and $\nabla^2 h_i$ are Lipschitz-continuous on $B_\delta(\bar{x})$, then the convergence is even q-quadratic.

Proof. We want to apply Theorem 10.5. This is possible since Lemma 18.2 guarantees that $F'(\bar{x}, \bar{\mu})$ is invertible.
For the q-quadratic convergence, one assumption of Theorem 10.5 is left to show, namely that F' is Lipschitz-continuous on $B_\delta(\bar{x}, \bar{\mu})$:
Applying the triangle inequality and using $\|A\| = \|A^\top\|$ leads to

$$\|F'(x, \mu) - F'(x', \mu')\| \leq \|\nabla^2_{xx} L(x, \mu) - \nabla^2_{xx} L(x', \mu')\| + 2\|\nabla h(x) - \nabla h(x')\|.$$

Due to Lemma 11.2 and since ∇h is continuously differentiable, it holds that ∇h is Lipschitz-continuous on $B_\delta(\bar{x}, \bar{\mu})$.
Moreover, it follows that

$$\|\nabla^2_{xx} L(x, \mu) - \nabla^2_{xx} L(x', \mu')\| \leq \|\nabla^2 f(x) - \nabla^2 f(x')\| \sum_{i=1}^{p} \|\mu_i \nabla^2 h_i(x) - \mu_i' \nabla^2 h_i(x')\|$$

and

$$\|\mu_i \nabla^2 h_i(x) - \mu_i' \nabla^2 h_i(x')\| \leq |\mu_i| \cdot \|\nabla^2 h_i(x) - \nabla^2 h_i(x')\| + |\mu_i - \mu_i'| \cdot \|\nabla^2 h_i(x')\|.$$

Hence, the Lipschitz-continuity of F' on $B_\delta(\bar{x}, \bar{\mu})$ follows since $\nabla^2 f$ and $\nabla^2 h_i$ are Lipschitz-continuous on $B_\delta(\bar{x})$. ∎

Local SQP Method

The Lagrange-Newton equation (18.4) can also be interpreted as the KKT conditions of the following quadratic optimization problem:

$$\begin{aligned} \min \ & q_k(s) = \nabla f(x^k)^\top s + \frac{1}{2} s^\top H_k s \\ \text{s.t.} \ & h(x^k) + \nabla h(x^k)^\top s = 0, \end{aligned} \tag{18.6}$$

where $H_k = \nabla^2_{xx} L(x^k, \mu^k)$.
Since all constraints of (18.6) are equalities and affine, every feasible point s of (18.6) fulfills a constraint qualification (see Theorem 15.26). Hence, every local optimum s^k of (18.6) fulfills the following KKT conditions:

$$\begin{aligned} \exists \mu^k_{qp} : \nabla f(x^k) + H_k s^k + \nabla h(x^k) \mu^k_{qp} &= 0 \\ h(x^k) + \nabla h(x^k)^\top s^k &= 0. \end{aligned}$$

Let $d^k_x = s^k$ and $d^k_\mu = \mu^k_{qp} - \mu^k$. Then it holds:

$$\begin{aligned} H_k d^k_x + \nabla h(x^k) d^k_\mu &= -\nabla f(x^k) - \nabla h(x^k) \mu^k \\ \nabla h(x^k)^\top d^k_x &= -h(x^k) \end{aligned}$$

This system is the Lagrange-Newton equation (18.4), i. e. $d^k = \begin{pmatrix} d^k_x \\ d^k_\mu \end{pmatrix}$ is a solution of (18.4).
Vice versa, if d^k is a solution of the Lagrange-Newton equation (18.4), then the vector $\begin{pmatrix} s^k \\ \mu^k_{qp} \end{pmatrix}$ with $s^k = d^k_x$ and $\mu^k_{qp} = d^k_\mu + \mu^k$ is a KKT pair of the quadratic optimization problem (18.6).
Thus, we get the following Lemma:

Lemma 18.4 *Let f and h be two-times continuously differentiable. Let $x^k \in \mathbb{R}^n$ and $\mu^k \in \mathbb{R}^p$. Then it holds:*
$d^k = \begin{pmatrix} d^k_x \\ d^k_\mu \end{pmatrix}$ solves the Lagrange-Newton equation (18.4) if and only if $(s^k, \mu^k_{qp}) = (d^k_x, \mu^k + d^k_\mu)$ is a KKT pair of the quadratic optimization problem (18.6).

Using the solution s^k of the SQP subproblem (18.6) and the corresponding multiplier, one can check whether x^k is a local optimum of (18.1):

Theorem 18.5 *Let f and h be two-times continuously differentiable. Let $x^k \in \mathbb{R}^n$ and $\mu^k \in \mathbb{R}^p$. Then the following statements are equivalent:*

a) (x^k, μ^k) is a KKT pair of (18.1) which satisfies the second-order sufficient conditions.

b) $s^k = 0$ is an isolated local optimum of (18.6) and $\mu^k_{qp} = \mu^k$ is the corresponding Lagrangian multiplier.

Proof.
"$a) \Rightarrow b)$": Since x^k is feasible for (18.1) it holds that $h(x^k) = 0$. Hence, it follows that

$$h(x^k) + \nabla h(x^k)^\top 0 = h(x^k) = 0,$$

i. e. $s^k = 0$ is feasible for (18.6). The Lagrange function of (18.6) is given by

$$L_k^{qp}(s, \mu_{qp}) = q_k(s) + (\mu_{qp})^\top (h(x^k) + \nabla h(x^k)^\top s).$$

Thus, we get that

$$\begin{aligned} \nabla_s L_k^{qp}(s, \mu_{qp}) &= \nabla f(x^k) + H_k s + \nabla h(x^k)\mu_{qp} \\ &= \nabla f(x^k) + \nabla h(x^k)\mu_{qp} + H_k s \\ &= \nabla_x L(x^k, \mu_{qp}) + H_k s. \end{aligned}$$

and that

$$\nabla_s L_{ss}^2 L_k^{qp}(s, \mu_{qp}) = H_k = \nabla_{xx}^2 L(x^k, \mu^k).$$

Hence, for $(s^k, \mu_{qp}^k) = (0, \mu^k)$, it is due to a) that

$$0 = \nabla_s L_k^{qp}(0, \mu^k) = \nabla_x L(x^k, \mu^k) \text{ and } s^\top H_k s > 0 \quad \forall s \in \mathbb{R}^n \setminus \{0\} \text{ with } \nabla h(x^k)^\top s = 0.$$

Thus, in $(s^k, \mu_{qp}^k) = (0, \mu^k)$, the second-order sufficient conditions of (18.6) hold, s^k is an isolated local minimum of (18.6) and $\mu_{qp}^k = \mu^k$ is the corresponding multiplier.

"$b) \Rightarrow a)$": Since $s^k = 0$ is feasible for (18.6), it follows that $0 = h(x^k) + \nabla h(x^k)^\top s^k = h(x^k)$. Together with the Lagrangian Multiplier $\mu_{qp}^k = \mu^k$, it follows that

$$0 = \nabla q_k(0) + \nabla h(x^k)\mu^k = \nabla f(x^k) + \nabla h(x^k)\mu^k = \nabla_x L(x^k, \mu^k).$$

Hence, (x^k, μ^k) is a KKT pair of (18.1).
Let $s \in \ker(\nabla h(x^k)) \setminus \{0\}$. Then, ts is feasible for (18.6) for all $t \in \mathbb{R}$.
Since $s^k = 0$ is an isolated local minimum of (18.6), it follows that $t = 0$ is an isolated minimum of the quadratic function $\varphi(t) = q_k(ts)$ and

$$0 < \varphi''(0) = s^\top \nabla^2 q_k(0) s = s^\top H_k s = s^\top \nabla_{xx}^2 L(x^k, \mu^k) s.$$

Hence, (x^k, μ^k) is a KKT pair satisfying the seconder-order sufficient conditions and x^k is an isolated local minimum of (18.1). ∎

This approach based on (18.6) leads to the following algorithm:

Algorithm 18.6 Local SQP method for (18.1)

1: Choose $x^0 \in \mathbb{R}^n$ and $\mu^0 \in \mathbb{R}^p$
2: **for** $k = 0, 1, 2, \ldots$ **do**
3: **if** (x^k, μ^k) is a KKT pair of (18.1) **then**
4: Stop
5: Compute optimal solution s^k of (18.6) with corresponding multiplier μ_{qp}^k
6: $x^{k+1} := x^k + s^k$; $\mu^{k+1} := \mu_{pq}^k$

SQP method for general constrained NLPs

Next, we consider the general constrained nonlinear program (15.1), which can have equalities and inequalities as constraints. Similarly to the equality-constrained case, we define the following quadratic SQP subproblem:

$$\begin{aligned} \min\ & q_k(s) = \nabla f(x^k)^\top s + \frac{1}{2} s^\top H_k s \\ \text{s.t.}\ & g(x^k) + \nabla g(x^k)^\top s \leq 0 \\ & h(x^k) + \nabla h(x^k)^\top s = 0. \end{aligned} \tag{18.7}$$

By generalizing Algorithm 18.6, we obtain the following algorithm for general constrained problems:

Algorithm 18.7 Local SQP method for (15.1)

1: Choose $x^0 \in \mathbb{R}^n$, $\lambda^0 \in \mathbb{R}^m$ and $\mu^0 \in \mathbb{R}^p$
2: **for** $k = 0, 1, 2, \ldots$ **do**
3: **if** (x^k, λ^k, μ^k) is a KKT triple of (15.1) **then**
4: Stop
5: Compute optimal solution s^k of (18.7) with corresponding multipliers λ^k_{qp} and μ^k_{qp}
6: $x^{k+1} := x^k + s^k$; $\lambda^{k+1} := \lambda^k_{qp}$; $\mu^{k+1} := \mu^k_{pq}$

Under suitable assumptions, one can get fast local convergence for this algorithm:

Theorem 18.8 *If*

a) f, g *and* h *are two-times continuously differentiable*

b) H_k *is chosen as* $\nabla^2_{xx} L(x^k, \lambda^k, \mu^k)$ *(i. e. exactly)*

c) $(\bar{x}, \bar{\lambda}, \bar{\mu})$ *is a KKT triple of (15.1)*

d) $\forall i \in U : g_i(\bar{x}) = 0 \Rightarrow \bar{\lambda}_i > 0$ *(strict complementarity)*

e) $(\nabla g_{\mathcal{A}(\bar{x})}(\bar{x}), \nabla h(\bar{x}))$ *has full column rank*

f) Second-order sufficient optimality conditions are satisfied:

$$s^\top \nabla^2_{xx} L(\bar{x}, \bar{\lambda}, \bar{\mu}) s > 0 \ \forall s \neq 0 \text{ with } \nabla g_{\mathcal{A}(\bar{x})}(\bar{x})^\top s = 0 \text{ and } \nabla h(\bar{x}) s = 0$$

g) among all KKT triples $(s^k, \lambda^k_{qp}, \mu^k_{qp})$ *of (18.7), the one with smallest distance*

$$\|(x^k + s^k, \lambda^k_{qp}, \mu^k_{qp}) - (x^k, \lambda^k, \mu^k)\|$$

is chosen in Algorithm 18.7,

then there is a $\delta > 0$ *such that for any* $(x^0, \lambda^0, \mu^0) \in B_\delta(\bar{x}, \bar{\lambda}, \bar{\mu})$*, Algorithm 18.7 either terminates or it generates a sequence* (x^k, λ^k, μ^k) *which converges q-superlinearly to* $(\bar{x}, \bar{\lambda}, \bar{\mu})$*.*
If, additionally, $\nabla^2 f$, $\nabla^2 g_i$ *and* $\nabla^2 h_i$ *are Lipschitz-continuous on* $B_\delta(\bar{x})$*, then the convergence is even q-quadratic.*

Globalized SQP method

Next, we want to consider a globalized SQP method, i. e. a method with global convergence. Here, we allow to choose an approximated matrix $H_k = H_k^\top$ in (18.7), e. g. by Quasi-Newton approximation (see chapter 13).
Moreover, we use the exact penalty function

$$P_\alpha^1(x) = f(x) + \alpha(\|(g(x))_+\|_1 + \|h(x)\|_1)$$

and the Armijo step length rule. Since P_α^1 may not necessarily be differentiable everywhere, we use directional derivatives for the Armijo step length rule:

Definition 18.9 Let $\varphi : \mathbb{R}^n \to \mathbb{R}$ be continuous.

$$D_+\varphi(x,s) = \lim_{t\to 0^+} \frac{\varphi(x+ts)-\varphi(x)}{t} \in \mathbb{R}$$

is called **directional derivative of φ in $\mathbf{x} \in \mathbb{R}^n$ along direction $\mathbf{s} \in \mathbb{R}^n$** (provided that the limit exists).

Theorem 18.10 *Let f, g and h be continuously differentiable and $\alpha > 0$. Then for any $x \in \mathbb{R}^n$ and any direction $s \in \mathbb{R}^n$ the directional derivative of P_α^1 in x along s exists and it is*

$$\begin{aligned} D_+P_\alpha^1(x,s) = \nabla f(x)^\top s + \alpha \sum_{g_i(x)>0} \nabla g_i(x)^\top s + \alpha \sum_{g_i(x)=0} (\nabla g_i(x)^\top s)_+ \\ +\alpha \sum_{h_i(x)>0} \nabla h_i(x)^\top s - \alpha \sum_{h_i(x)<0} \nabla h_i(x)^\top s + \sum_{h_i(x)=0} |\nabla h_i(x)^\top s| \end{aligned}$$

Proof. We treat each term separately:
It is

$$\lim_{t\to 0^+} \frac{f(x+ts)-f(x)}{t} = \nabla f(x)^\top s.$$

If $g_i(x) > 0$, then $(g_i)_+ = g_i$ close to x and

$$\lim_{t\to 0^+} \frac{(g_i(x+ts))_+ - (g_i(x))_+}{t} = \lim_{t\to 0^+} \frac{g_i(x+ts)-g_i(x)}{t} = \nabla g_i(x)^\top s.$$

If $g_i(x) < 0$, then $(g_i)_+ = 0$ close to x and

$$\lim_{t\to 0^+} \frac{(g_i(x+ts))_+ - (g_i(x))_+}{t} = 0.$$

If $g_i(x) = 0$, then $g_i(x+ts) = t\nabla g_i(x)^\top s + o(t)$ and

$$\lim_{t\to 0^+} \frac{(g_i(x+ts))_+ - (g_i(x))_+}{t} = \lim_{t\to 0^+} (\nabla g_i(x)^\top s + \frac{o(t)}{t})_+ = (\nabla g_i(x)^\top s)_+.$$

If $h_i(x) > 0$, then $|h_i| = h_i$ close to x and

$$\lim_{t\to 0^+} \frac{|h_i(x+ts)| - |h_i(x)|}{t} = \nabla h_i(x)^\top s.$$

If $h_i(x) < 0$, then $|h_i| = -h_i$ close to x and

$$\lim_{t\to 0^+} \frac{|h_i(x+ts)| - |h_i(x)|}{t} = -\nabla h_i(x)^\top s.$$

If $h_i(x) = 0$, then

$$\lim_{t\to 0^+} \frac{|h_i(x+ts)| - |h_i(x)|}{t} = \lim_{t\to 0^+} |\nabla h_i(x)^\top s + \frac{o(t)}{t}| = |\nabla h_i(x)^\top s|. \qquad \blacksquare$$

The next theorem shows that if the penalty parameter is large enough and H_k is positive definite, then every KKT point s^k of (18.7) defines a descent direction of P^1_α:

Theorem 18.11 *Let f, g and h be continuously differentiable and let $(s^k, \lambda^k_{qp}, \mu^k_{qp})$ be a KKT triple of (18.7). Let*

$$\alpha \geq \max\{(\lambda^k_{qp})_1, \ldots, (\lambda^k_{qp})_m, |(\mu^k_{qp})_1|, \ldots, |(\mu^k_{qp})_p|\}.$$

Then it holds that

$$D_+P^1_\alpha(x^k, s^k) \leq -{s^k}^\top H_k s^k.$$

If H_k is positive definite, then s^k is a descent direction for P^1_α.

Proof. Due to the complementarity condition, it holds that

$$(\lambda^k_{qp})_i \geq 0, \quad g_i(x^k) + \nabla g_i(x^k)^\top s^k \leq 0 \quad (\lambda^k_{qp})_i(g_i(x^k) + \nabla g_i(x^k)^\top s^k) = 0.$$

Thus, it follows that

$$\begin{aligned}(\lambda^k_{qp})^\top \nabla g(x^k)^\top s^k &= \sum_{g_i(x^k)>0} (\lambda^k_{qp})_i \nabla g_i(x^k)^\top s^k - \sum_{g_i(x)\leq 0} \underbrace{(\lambda^k_{qp})_i}_{\geq 0} \underbrace{g_i(x^k)}_{\leq 0} \\ &\geq \sum_{g_i(x^k)>0} (\lambda^k_{qp})_i \underbrace{\nabla g_i(x^k)^\top s^k}_{<0} \geq \alpha \sum_{g_i(x^k)>0} \nabla g_i(x^k)^\top s^k.\end{aligned}$$

Moreover, $h_i(x^k) + \nabla h_i(x^k)^\top s^k = 0$ leads to

$$\begin{aligned}{\mu^k_{qp}}^\top \nabla h(x^k)^\top s^k &= \sum_{h_i(x^k)>0} (\mu^k_{qp})_i \underbrace{\nabla h_i(x^k)^\top s^k}_{<0} + \sum_{h_i(x^k)<0} (\mu^k_{qp})_i \underbrace{\nabla h_i(x^k)^\top s^k}_{>0} \\ &\geq \alpha \sum_{h_i(x^k)>0} \nabla h_i(x^k)^\top s^k - \alpha \sum_{h_i(x^k)<0} \nabla h_i(x^k)^\top s^k.\end{aligned}$$

The multiplicator rule says that

$$\nabla f(x^k) + H_k s^k + \nabla g(x^k)\lambda^k_{qp} + \nabla h(x^k)\mu^k_{qp} = 0.$$

Thus, it follows that

$$\begin{aligned}\nabla f(x^k)^\top s^k &= -{s^k}^\top H_k s^k - {\lambda_{qp}^k}^\top \nabla g(x^k)^\top s^k - {\mu_{qp}^k}^\top \nabla h(x^k)^\top s^k \\ &\leq -{s^k}^\top H_k s^k - \alpha \sum_{g_i(x^k)>0} \nabla g_i(x^k)^\top s^k - \alpha \sum_{h_i(x^k)>0} \nabla h_i(x^k)^\top s^k \\ &\quad + \alpha \sum_{h_i(x^k)<0} \nabla h_i(x^k)^\top s^k\end{aligned}$$

and that

$$\begin{aligned}D_+ P_\alpha^1(x^k, s^k) \overset{\text{Thm. } 18.10}{=} &\nabla f(x^k)^\top s^k + \alpha \sum_{g_i(x^k)>0} \nabla g_i(x^k)^\top s^k + \alpha \sum_{g_i(x^k)=0} (\nabla g_i(x^k)^\top s^k)_+ \\ &+ \alpha \sum_{h_i(x^k)>0} \nabla h_i(x^k)^\top s^k - \alpha \sum_{h_i(x^k)<0} \nabla h_i(x^k)^\top s^k \\ &+ \alpha \sum_{h_i(x^k)=0} |\nabla h_i(x^k)^\top s^k| \qquad \blacksquare \\ \leq &- {s^k}^\top H_k s^k + \alpha \sum_{g_i(x^k)=0} (\underbrace{\nabla g_i(x^k)^\top s^k}_{\leq -g_i(x^k)=0})_+ + \alpha \sum_{h_i(x^k)=0} |\underbrace{\nabla h_i(x^k)^\top s^k}_{=-h_i(x^k)=0}| \\ = &- {s^k}^\top H_k s^k.\end{aligned}$$

One globalized SQP method is then given as follows:

Algorithm 18.12 Globalized SQP method for (15.1)

1: Choose $x^0 \in \mathbb{R}^n$, $\lambda^0 \in \mathbb{R}^m$, $\mu^0 \in \mathbb{R}^p$, $H_0 \in \mathbb{R}^{n\times n}$ symmetric, $\alpha > 0$ sufficiently large and $0 < \gamma < \frac{1}{2}$
2: **for** $k = 0, 1, 2, \ldots$ **do**
3: **if** (x^k, λ^k, μ^k) is a KKT triple of (15.1) **then**
4: Stop
5: Compute solution s^k of (18.7) with corresponding multipliers λ_{qp}^k and μ_{qp}^k
6: Compute the largest number $\sigma_k \in \{1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots\}$ satisfying

$$P_\alpha^1(x^k + \sigma_k s^k) - P_\alpha^1(x^k) \leq \gamma \sigma_k D_+ P_\alpha^1(x^k, s^k).$$

7: $x^{k+1} := x^k + \sigma_k s^k$, compute new multipliers λ^{k+1}, μ^{k+1} and choose a new symmetric matrix $H_{k+1} \in \mathbb{R}^{n\times n}$.

Potential practical problems

A problem in practice is that some SQP subproblems might not be solvable or even infeasible:

Example 18.13 We consider the problem

$$\begin{aligned} \min\ & f(x) \\ \text{s.t.}\ & g(x) := 1 - x^2 \leq 0 \\ & x \in \mathbb{R}, \end{aligned}$$

In $x^k = 0$, it is $\nabla g(x^k) = 0$ and

$$g(x^k) + \nabla g(x^k)^\top s = 1 \not\leq 0 \quad \forall s \in \mathbb{R}.$$

Thus, the constraint of this subproblem is never satisfied.
One idea to handle such problems is to relax the feasibility and, instead of (18.7), solve the relaxed problem

$$\begin{aligned} \min\ & \nabla f(x^k)^\top s + \frac{1}{2} s^\top H_k s + \rho \sum_{i=1}^{m} v_i + \rho \sum_{i=1}^{p} ((w_+)_i + (w_-)_i) \\ \text{s.t.}\ & g(x^k) + \nabla g(x^k)^\top s \leq v \\ & h(x^k) + \nabla h(x^k)^\top s = w_+ - w_- \\ & v \geq 0 \\ & w_+, w_- \geq 0 \\ & s \in \mathbb{R}^n, v \in \mathbb{R}^m, w_+, w_- \in \mathbb{R}^p \end{aligned} \tag{18.8}$$

The vectors v, w_+ and w_- allow the violation of the original constraints of the SQP subproblem (18.7), but penalize the violation in the objective function. The variable ρ is a penalty parameter. It is clear that the relaxed SQP subproblem (18.8) has feasible solutions. One can show the following connection between the SQP subproblem (18.7) and the relaxed problem (18.8):

Theorem 18.14

1. *If $(s^k, \lambda_{qp}^k, \mu_{qp}^k)$ is a KKT triple of (18.7), then for every*

   $$\rho \geq \max\{(\lambda_{qp}^k)_1, \ldots, (\lambda_{qp}^k)_m, |(\mu_{qp}^k)_1|, \ldots, |(\mu_{qp}^k)_p|\}$$

   the vector $(s^k, v^k, w_+^k, w_-^k, \lambda_{qp}^k, \mu_{qp}^k, \xi^k, \xi_+^k, \xi_-^k)$ with

   $$v^k = 0, \quad w_+^k = 0, \quad w_-^k = 0, \quad \xi^k = \rho e - \lambda_{qp}^k, \quad \xi_+^k = \rho e - \mu_{qp}^k, \quad \xi_-^k = \rho e + \mu_{qp}^k$$

   is a KKT triple of (18.8). ($e = (1, \ldots, 1)^\top$, ξ^k are the Lagrangian multipliers to $-v \leq 0$, ξ_+^k the ones to $-w_+ \leq 0$ and ξ_-^k the ones to $-w_- \leq 0$.)

2. *If $(s^k, 0, 0, 0, \lambda_{qp}^k, \mu_{qp}^k, \xi^k, \xi_+^k, \xi_-^k)$ is a KKT triple of (18.8), then $(s^k, \lambda_{qp}^k, \mu_{qp}^k)$ is a KKT triple of (18.7).*

19. Quadratic Optimization Problems

In chapter 18, we need to solve SQP subproblems, which are quadratic optimization roblems. In this chapter, we consider such problems

$$\begin{aligned} \min\ & q(x) := c^\top x + \frac{1}{2} x^\top H x \\ \text{s.t.}\ & g(x) := A^\top x + \alpha \leq 0 \\ & h(x) := B^\top x + \beta = 0, \end{aligned} \tag{QP}$$

where $c \in \mathbb{R}^n$, $H = H^\top \in \mathbb{R}^{n \times n}$, $A \in \mathbb{R}^{n \times m}$, $\alpha \in \mathbb{R}^m$, $B \in \mathbb{R}^{n \times p}$ and $\beta \in \mathbb{R}^p$.

If the quadratic problem (QP) is non-convex, then there might be many isolated solutions:

Example 19.1 For the quadratic problem

$$\begin{aligned} \min\ & -\frac{1}{6} \sum_{l=1}^{n} 2^l ((x_l - l)^2 - 1) \\ \text{s.t.}\ & 0 \leq x \leq 3, \end{aligned}$$

every of the 2^n vertices of the feasible set $[0,3]^n$ is an isolated local minimum and all these points have different objective function values: $0, -1, -2, \ldots, -2^n + 1$.

In the following, we will therefore only consider strictly convex quadratic problems, i. e. we assume that H is positive definite. In this case, it follows from Theorem 15.32 and the fulfilled constraint qualification (15.4) that $\bar{x}$ is an optimal solution of (QP) if and only if $\bar{x}$ is a KKT point of (QP).
If (QP) has a feasible solution $\hat{x}$, then $N_q(\hat{x}) = \{x \in \mathbb{R}^n : q(x) \leq q(\hat{x})\}$ is compact because of $q(x) \overset{\|x\| \to \infty}{\longrightarrow} \infty$. Then $N_q(\hat{x}) \cap X \neq \emptyset$ is also compact and q has a global minimum on $N_q(\hat{x}) \cap X$, which is also the optimal solution of (QP). From Theorem 6.6 and the convexity of the feasible set of (QP), it follows that this optimal solution is even unique.
Hence, it holds that (QP) has an optimal solution if and only (QP) has a feasible solution.

First, we consider quadratic problems (QP) without inequality constraints:
In this case $\bar{x}$ solves (QP) if and only if there exists some $\bar{\mu} \in \mathbb{R}^p$ such that the KKT conditions

$$\begin{pmatrix} H & B \\ B^\top & 0 \end{pmatrix} \begin{pmatrix} \bar{x} \\ \bar{\mu} \end{pmatrix} = \begin{pmatrix} -c \\ -\beta \end{pmatrix} \tag{19.1}$$

hold.
If $X \neq \emptyset$, then (19.1) has a solution.
If $\operatorname{rank}(B) = p$, then (19.1) has a unique solution and vice versa.
Thus, the quadratic problem (QP) without inequality constraints is equivalent to solving a linear system of equations and therefore easy to solve.

Next, we consider general strictly convex quadratic problems (QP):
Our idea is to solve this problem by solving a sequence of quadratic problems without inequality constraints. More precisely, when computing x^{k+1}, we approximate the set of active constraints $\mathcal{A}(x^k)$ in x^k "from inside", i. e. we choose $\mathcal{A}_k \subseteq \mathcal{A}(x^k)$ and treat these constraints like equality constraints while ignoring the remaining inequality constraints. Thus, we solve in each iteration a problem

$$\begin{aligned} &\min\ q(x) \\ &\text{s.t.}\ \ A_{\mathcal{A}_k}^\top x + \alpha_{\mathcal{A}_k} = 0, \\ &\qquad B^\top x + \beta = 0, \end{aligned} \tag{QP_k}$$

with A_I being the columns a_i, $i \in I \subseteq \{1, \dots, m\}$ of the matrix $A = (a_1, \dots, a_m)$.
This leads to the following algorithm:

Algorithm 19.2 Active sets strategy

1: Choose a starting point $x^0 \in \mathbb{R}^n$ that is feasible for (QP), set $\mathcal{A}_0 := \mathcal{A}(x^0)$
2: **for** $k = 0, 1, 2, \dots$ **do**
3: $\quad \mathcal{I}_k := U \setminus \mathcal{A}_k$, $\lambda_{\mathcal{I}_k}^{k+1} := 0$ and compute a KKT triple $(\hat{x}^{k+1}, \lambda_{\mathcal{A}_k}^{k+1}, \mu^{k+1})$ of (QP_k)
$\quad d^k := \hat{x}^{k+1} - x^k$
4: $\quad$ **if** $d^k = 0$ and $\lambda^{k+1} \geq 0$ **then**
5: $\quad\quad x^{k+1} := x^k$ and stop; return KKT triple $(x^{k+1}, \lambda^{k+1}, \mu^{k+1})$ of (QP).
6: $\quad$ **if** $d^k = 0$ and there exists some $j \in \mathcal{A}_k$ with $\lambda_j^{k+1} = \min\limits_{i \in \mathcal{A}_k} \lambda_i^{k+1} < 0$ **then**
7: $\quad\quad x^{k+1} := x^k$, $\mathcal{A}_{k+1} := \mathcal{A}_k \setminus \{j\}$ and goto next iteration
8: $\quad$ **if** $d^k \neq 0$ and $\hat{x}^{k+1}$ is feasible for (QP) **then**
9: $\quad\quad x^{k+1} := \hat{x}^{k+1}$, $\mathcal{A}_{k+1} := \mathcal{A}_k$ and goto next iteration.
10: $\quad$ **if** $d^k \neq 0$ and $\hat{x}^{k+1}$ is not feasible for (QP) **then**
11: $\quad\quad$ compute

$$\sigma_k = \max\{\sigma \geq 0 : x^k + \sigma d^k \text{ feasible for (QP)}\}$$

$\quad\quad$ and an index $j \in \mathcal{I}_k$ with $a_j^\top (x^k + \sigma_k d^k) + \alpha_j = 0$.
$\quad\quad$ Set $x^{k+1} := x^k + \sigma_k d^k$ and $\mathcal{A}_{k+1} := \mathcal{A}_k \cup \{j\}$.

Theorem 19.3 *Consider Algorithm 19.2. It holds:*

a) x^k is feasible for (QP_k) and (QP).

b) If $d^k \neq 0$ and $\hat{x}^{k+1}$ is not feasible for (QP), then the step length σ_k in line 11 exists as well as index j and it holds $0 \leq \sigma_k < 1$,

$$\sigma_k = \min\left\{ -\frac{{a_i}^\top x^k + \alpha_i}{{a_i}^\top d^k} : i \in \mathcal{I}_k, {a_i}^\top d^k > 0 \right\}$$

c) *If $d^k = 0$ and $\lambda^{k+1} \geq 0$, then $(x^{k+1}, \lambda^{k+1}, \mu^{k+1})$ is a KKT triple of (QP).*

d) *If $d^k \neq 0$, then $\nabla q(x^k)^\top d^k < 0$, i. e. d^k is a descent direction for q in x^k. Moreover, $q(x^{k+1}) < q(x^k)$, if $x^{k+1} \neq x^k$.*

e) *For every x^k there exists some $l \geq k$ such that x^l is the unique global optimal solution of (QP_l).*

f) *If Algorithm 19.2 does not terminate after a finite number of steps, then there is some $l \geq 0$ with $x^k = x^l$ for all $k \geq l$.*

g) *If the columns of the matrix $(A_{\mathcal{A}_k}, B)$ are linearly independent, then the columns of the matrix $(A_{\mathcal{A}_{k+1}}, B)$ are also linearly independent.*

Proof.

a) Since x^0 is feasible for (QP) and $\mathcal{A}_0 = \mathcal{A}(x^0)$, it is feasible for (QP_0).
Let x^k be feasible for (QP_k) and (QP).
The vector $\hat{x}^{k+1}$ is feasible for (QP_k).
Since x^{k+1} is located on the line segment between x^k and $\hat{x}^{k+1}$ and the feasible set of (QP_k) is convex, it follows that x^{k+1} is feasible for (QP_k).
It is $\mathcal{A}_{k+1} \subseteq \mathcal{A}_k$ (lines 6-7 & lines 8-9) or $\mathcal{A}_{k+1} = \mathcal{A}_k \cup \{j\}$ (lines 10-11).
In both cases, x^{k+1} is feasible for (QP_{k+1}).
If $x^{k+1} = x^k$ (lines 4-5 & lines 6-7), x^{k+1} is feasible for (QP).
If $x^{k+1} = \hat{x}^{k+1} \neq x^k$ (lines 8-9), then $\hat{x}^{k+1} (= x^{k+1})$ is feasible for (QP).
If $x^{k+1} = x^k + \sigma_k d^k$, then the choice of σ_k is such that x^{k+1} is feasible for (QP).

b) If $\hat{x}^{k+1}$ is not feasible for (QP), but feasible for (QP_k), then there must be some constraint with index $i \in \mathcal{I}_k$ which is violated, i. e.

$$a_i^\top (x^k + d^k) + \alpha_i > 0.$$

Because of a) it is $a_i^\top x^k + \alpha_i \leq 0$, i. e. it must hold $a_i^\top d^k > 0$.
For indices i with $a_i^\top d^k \leq 0$, $x^k + \sigma d^k$ is feasible for all $\sigma \in [0, 1]$.
Hence, σ_k is the largest number such that

$$a_i^\top (x^k + \sigma_k d^k) + \alpha_i \leq 0 \quad \forall i \in \mathcal{I}_k, a_i^\top d^k > 0$$

and b) follows.

c) Let $d^k = 0$ and $\lambda^{k+1} \geq 0$.
It follows that $x^{k+1} = x^k$, $\lambda^{k+1}_{\mathcal{I}_k} = 0$ and that $(x^{k+1}, \lambda^{k+1}_{\mathcal{A}_k}, \mu^{k+1})$ is a KKT triple of (QP_k).
From the multiplicator rule, it follows that

$$\nabla q(x^{k+1}) + A\lambda^{k+1} + B\mu^{k+1} = \nabla q(x^{k+1}) + A_{\mathcal{A}_k}\lambda^{k+1}_{\mathcal{A}_k} + B\mu^{k+1} = 0.$$

Moreover, x^{k+1} is feasible for (QP) and $\lambda^{k+1} \geq 0$ and

$$\lambda^{k+1^\top} g(x^{k+1}) = \lambda_{\mathcal{A}_k}^{k+1^\top} g_{\mathcal{A}_k}(x^{k+1}) = 0.$$

Hence, $(x^{k+1}, \lambda^{k+1}, \mu^{k+1})$ is a KKT triple of (QP).

d) if $d^k \neq 0$, then x^k does not coincide with the unique global optimum $\hat{x}^{k+1}$ of (QP_k). Since the feasible set of (QP_k) is convex and its objective function is strictly convex, it follows from Theorem 6.6 that $\hat{x}^{k+1}$ is the unique global optimum of (QP_k). Hence, it is $q(\hat{x}^{k+1}) < q(x^k)$ and, due to the convexity of q, it follows from Theorem 6.3 that

$$\nabla q(x^k)^\top d^k \leq q(\hat{x}^{k+1}) - q(x^k) < 0.$$

If $x^{k+1} \neq x^k$, then either $x^{k+1} := \hat{x}^{k+1}$ or $x^{k+1} = x^k + \sigma_k d^k$ with $\sigma_k \in (0,1)$. With $\tau_k = 1$ or $\tau_k = \sigma_k$ and the convexity of q, we get that

$$q(x^{k+1}) - q(x^k) \leq (1-\tau_k) q(x^k) + \tau_k q(\hat{x}^{k+1}) - q(x^k) = \tau_k \underbrace{(q(\hat{x}^{k+1}) - q(x^k))}_{<0} < 0.$$

e) Let x^k, $k \in \mathbb{N}$ be given. There can occur three different cases:
Case 1: $d^k = 0$: Then, $x^k = \hat{x}^{k+1}$ is the unique global minimum of (QP_k).
Case 2: $d^k \neq 0$, $x^{k+1} = \hat{x}^{k+1}$ and $\mathcal{A}_{k+1} = \mathcal{A}_k$:
Then x^{k+1} is the unique global minimum of (QP_{k+1}).
Case 3: $d^k \neq 0$, $x^{k+1} = x^k + \sigma_k d^k$ and $\mathcal{A}_{k+1} = \mathcal{A} \cup \{j\}$
Since U is finite, the third case can only occur finitely often (at least consecutively). Hence, after finitely many iterations, we find an x^l which is the unique optimal solution of (QP_l).

f) If the algorithm does not terminate, then, by e), there exists an infinite sequence $(l_i)_i$ where x^{l_i} is the unique optimum of (QP_{l_i}). Since the power set of U is finite, there must be an infinite subsequence $(l_i')_i \subset (l_i)_i$ with $\mathcal{A}_{l_i'} = \mathcal{A}_{l_1'}$ and therefore $x^{l_i'} = x^{l_1'}$ for all i. The existence of an index $l_i' \leq k < l_{i+1}'$ with $x^k \neq x^{k+1}$ would lead to the contradiction

$$q(x^{l_i'}) = q(x^{l_{i+1}'}) \leq q(x^{k+1}) < q(x^k) \leq q(x^{l_i'}).$$

Hence, it must hold that $x^k = x^{l_1'}$ for all $k \geq l_1'$.

g) The only relevant case for the proof is the case $\mathcal{A}_{k+1} \not\subseteq \mathcal{A}_k$.
Then it must hold that $d^k \neq 0$, $\mathcal{A}_{k+1} = \mathcal{A}_k \cup \{j\}$ and $a_j^\top d^k > 0$.
If $a_j = A_{\mathcal{A}_k} v + Bw$ for some suitable vectors v and w, then we would get the contradiction

$$0 < a_j^\top d^k = v^\top \underbrace{A_{\mathcal{A}_k}^\top d^k}_{=0} + w^\top \underbrace{B^\top d^k}_{=0} = 0. \qquad \blacksquare$$

The last theorem shows that Algorithm 19.2 terminates with an optimal solution or it remains in the same point after a certain number of iterations and does not terminate. The latter behaviour is called cycling. In practice, cycling occurs very rarely.

Bibliography

[1] Sven O. Krumke. Nonlinear optimization, lecture notes, 2011.

[2] M. Ulbrich and S. Ulbrich. *Nichtlineare Optimierung.* Mathematik Kompakt. Springer Basel, 2012.