If the function is twice differentiable and the Hessian is positive semidefinite in the entire domain, then the function is convex. Note that the domain must be assumed to be convex too. If the Hessian has a negative eigenvalue at a point in the interior of the domain, then the function is not convex.

Definition (Descent direction): Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function and $x \in \mathbb{R}^n$. A vector $s \in \mathbb{R}^n \setminus \{0\}$ is called a descent direction of f at x if the directional derivative is negative, i.e. $\partial_s f(x) = \nabla f(x)^T s < 0$.

Remark: Let us define the restriction of f on the ray $\{\sigma \in \mathbb{R}_0^+ : x + \sigma s\}$ as

$$\varphi : \begin{cases} \mathbb{R}_+ \to \mathbb{R} \\ \sigma \quad \mapsto \varphi(\sigma) := f(x + \sigma s) \end{cases} \tag{1}$$

then $\nabla f(x)^T s = \varphi'(0)$ hence

- $s \in \mathbb{R}^n$ is a descent direction if and only if $\varphi'(0) < 0$
- if s is a descent direction, then φ is strictly monotonously decreasing in a neighborhood of $\sigma = 0$ (note, the reverse is not true).

Lemma (Descent poperty of descent direction): Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function, $x \in \mathbb{R}^n$ and s a descent direction of f at x. Then there exists an $\varepsilon > 0$ such that

$$f(x + \sigma s) < f(x) \qquad \forall \sigma \in (0, \varepsilon] \tag{2}$$

Proof: Since $\partial_s f(x) < 0$, it follows from the definition of the directional derivative that, whne taking the right-handed limit only,

$$\partial_s f(x) = \lim_{\sigma \to 0^+} \frac{f(x + \sigma s) - f(x)}{\sigma} = \lim_{\sigma \to 0^+} F(\sigma) < 0. \tag{3}$$

By continuity of the limit process there is an $\varepsilon > 0$ such that at this value $F(\varepsilon) < 0$ holds and thus in particular for any $\sigma \in (0, \varepsilon]$ it holds that

$$\frac{f(x + \sigma s) - f(x)}{\sigma} < 0 \implies f(x + \sigma s) < f(x) \tag{4}$$

□

Constant step size $\sigma_k = \sigma$ for all k. The main advantage of teh constatnt step size strategy is of course its simplicity, but at this point it is unclear how to choose the constant. A large constant might cause the algorithm to be nondecreasing, and a small constant can cause slow onvergence of the method.

Exact line search σ_k is a minimizer of f along the ray $x^k + \sigma s^k$

$$\sigma_k \in \underset{\sigma \geq 0}{\operatorname{argmin}}\, f(x^k + \sigma s^k) \tag{5}$$

This seems attractive, from a first glance, but it is not always possible to actually find the exact minimizer.

Backtracking The method requires $S_0 > 0, \beta \in (0,1)$ and $\gamma \in (0,1)$. The choice of σ_k is done by the following procedure. Frist, set σ_k to the initial guess S_0. Then, while

$$f(x^k) - f(x^k + \sigma_k s^k) < -\gamma \sigma_k \nabla f(x^k)^T s^k \tag{6}$$

we set $\sigma_k = \beta \sigma_k$. In other words the step size is chosen as $\sigma_k = S_0 \beta^{i_k}$, where i_k is the smallest non-negative integer for which the condition

$$f(x^k) - f(x^k + S_0 \beta^{i_k} s^k) \geq -\gamma S_0 \beta^{i_k} f(x^k)^T s^k \tag{7}$$

is satisfied. This gives a sequence $(S_0, S_0\beta, S_0\beta^2, S_0\beta^3, \ldots)$.

Lemma (Validity of the sufficient decrease condition): Let $U \subseteq \mathbb{R}^n$ be open, $f \in \mathcal{C}^1(U, \mathbb{R}), x \in U$, and $\gamma \in (0,1)$. Suppose $s \in \mathbb{R}^n \setminus \{0\}$ is a descent direction of f at x. Then there is a $\varepsilon > 0$ such that

$$f(x) - f(x + \sigma s) \geq -\gamma \sigma \nabla f(x)^T s \qquad \forall \sigma \in [0, \varepsilon] \tag{8}$$

Proof: The inequality holds for $\sigma = 0$. Let $\sigma > 0$. Since f is continuously differentiable it follows with Talyor expansion that

$$f(x + \sigma s) = f(x) + \sigma \nabla f(x)^T s + o(\sigma \|s\|), \tag{9}$$

where $x + \sigma s \in U$ and hence

$$f(x) - f(x + \sigma s) = -\gamma \sigma \nabla f(x)^T s - (1 - \gamma)\sigma \nabla f(x)^T s - o(\sigma \|s\|). \tag{10}$$

Since s is a descent direction of f at x we have

$$\lim_{\sigma \to 0^+} \frac{(1-\gamma)\sigma \nabla f(x)^T s + o(\sigma \|s\|)}{\sigma} = (1 - \gamma) \nabla f(x)^T s < 0. \tag{11}$$

Hence, there is an $\varepsilon > 0$ such that for all $\sigma \in (0, \varepsilon]$ the inequality

$$(1-\gamma)\sigma \nabla f(x)^T s - o(\sigma \|s\|) < 0 \tag{12}$$

holds, which combined with Equation 10 implies the desired result. □

Definition (Steepest descent): Let $x \in \mathbb{R}^n$ and $\nabla f(x) \neq 0$. Let $\bar{d} \in \mathbb{R}^n$ be the solution of the steepest descent problem

$$\min_{d \in \mathbb{R}^n, \|d\| = 1} \nabla f(x)^T d \tag{13}$$

where $\|\cdot\|$ is the Euclidean norm (of course other norms can be considered as well). Every direction $s = \lambda \bar{d}$ with $\lambda > 0$ is called a direction of steepest descent with respect to the Euclidean norm.

Theorem (Solution of the steepest descent problem): Let $x \in \mathbb{R}^n$ and $\nabla f(x) \neq 0$. Then the steepest descent problem has the unique solution

$$\bar{d} = -\frac{\nabla f(x)}{\|\nabla f(x)\|}. \tag{14}$$

This means that every direction $s = -\lambda \nabla f(x), \lambda > 0$, is direction of steepest descent.

Example (1.2): Show, that for a continuous function $f : \mathbb{R}^n \to \mathbb{R}$ with $\lim_{\|x\| \to \infty} f(x) = \infty$ and for arbitrary $w \in \mathbb{R}^n$ the level set

$$\mathcal{N}_{f(w)} = \{x \in \mathbb{R}^n : f(x) \leq f(w)\} \tag{15}$$

is compact.

Solution:
- Closedness: implied by continuity of f.
- Boundedness: assume there is some w for which $\{x \in \mathbb{R}^n : f(x) \le f(w)\}$ is unbounded. Then there must be a sequence $\{x^k\}_{k\in\mathbb{N}_0} \subset S$ with $\|x^k\| \to \infty$ but since f is coercive, $\lim_{\|x\|\to\infty} f(x) = \infty$ which contradicts $f(x) \le f(w)$

Example (2.1.a): Let $f : \mathbb{R} \to \mathbb{R}$ be a differentiable function. Show: if $\bar{x}$ is a local but not a global minimu of f, then f posseses at least one additional stationary point $\hat{x}$ besides $\bar{x}$, i.e. at least one point $\hat{x} \neq \bar{x}$ with $f'(x) = 0$.

Solution: There exists some $\varepsilon > 0$ with $f(x) \ge f(\bar{x})$ for all $x \in B_\varepsilon(\bar{x})$ and some $\tilde{x} \in \mathbb{R}$ with $f(\tilde{x}) < f(\bar{x})$.

Example: Let $f(x) = \frac{1}{2}x^T C x + c^T x$ a strictly convex function where $C \in R^{n\times n}$ positive definite.

$$\varphi'(\sigma_k) = \left(c + C\left(x^k + \sigma_k s^k\right)\right) = 0. \tag{16}$$

we get

$$\begin{aligned}\sigma_k &= -\frac{\left(c + Cx^k\right)^T s^k}{\left(s^k\right)^T C s^k} = -\frac{\nabla f\left(x^k\right)^T s^k}{\left(s^k\right)^T C s^k} \\ &= \frac{\left\|\nabla f\left(x^k\right)\right\|^2}{\nabla f\left(x^k\right)^T C \nabla f\left(x^k\right)} = \frac{\left\|s^k\right\|^2}{\left(s^k\right)^T C s^k}\end{aligned} \tag{17}$$

Theorem: Let $f : \mathbb{R}^n \to \mathbb{R}$ be strictly convex and quadratic, i.e. $f(x) = \frac{1}{2}x^T C x + c^T x$. Let the sequences $\{x^k\}_{k\in\mathbb{N}_0}$ and $\{\sigma_k\}_{k\in\mathbb{N}_0}$ be generated bt the gradient descent method with exact line search. Then it holds that

$$f\left(x^{k+1}\right) - f(\bar{x}) \le \left(\frac{\lambda_{\max}(C) - \lambda_{\min}(C)}{\lambda_{\max}(C) - \lambda_{\max}(C)}\right)^2 \cdot \left(f\left(x^k\right) - f(\bar{x})\right) \tag{18}$$

$$\left\|x^k - \bar{x}\right\| \le \sqrt{\frac{\lambda_{\max}(C)}{\lambda_{\min}(C)}} \cdot \left(\frac{\lambda_{\max}(C) - \lambda_{\min}(C)}{\lambda_{\max}(C) + \lambda_{\min}(C)}\right)^k \cdot \left\|x^0 - \bar{x}\right\| \tag{19}$$

where $\bar{x} = -C^{-1}c$ is the global minimum of f and $\lambda_{\max}(C)$ and $\lambda_{\min}(C)$ are the larges and smallest eigenvalue of C, respectively. With the condition $\kappa(C) = \frac{\lambda_{\max}(C)}{\lambda_{\min}(C)}$ of the positive definite matrix C the previous two inequalities can be expressed in the equivalent form:

$$f\left(x^{k+1}\right) - f(\bar{x}) \le \left(\frac{\kappa(C) - 1}{\kappa(C) + 1}\right)^2 \cdot \left(f\left(x^k\right) - f(\bar{x})\right), \tag{20}$$

$$\left\|x^k - \bar{x}\right\| \le \sqrt{\kappa(C)} \cdot \left(\frac{\kappa(C) - 1}{\kappa(C) + 1}\right)^k \cdot \left\|x^0 - \bar{x}\right\|. \tag{21}$$