it follows that

$$S^k + \nu^k = -\left(Q_k + \frac{1}{t_k}\mathbb{I}\right) d^k, \tag{2.4}$$

S^k and ν^k being the augmented aggregate subgradient and outer normal defined in (1.4) and (1.8) respectively.

From this the model decrease (1.7) can be calculated using (1.6), (1.9) and (2.4):

$$\begin{aligned}
\delta_k &:= \hat{f}_k - M_k(x^{k+1}) - \left\langle \nu^k, d^k \right\rangle \\
&= \hat{f}_k - A_k(x^{k+1}) - \left\langle \nu^k, d^k \right\rangle \\
&= C_k - \left\langle S^k + \nu^k, d^k \right\rangle \\
&= C_k + \left\langle d^k, \left(Q_k + \frac{1}{t_k}\mathbb{I}\right) d^k \right\rangle .
\end{aligned} \tag{2.5}$$

The new δ_k is used in the same way as in algorithm 1.1 for the descent test and stopping conditions.

Because the changes in the algorithm concern only the stabilization and the decrease measure δ_k all other relations that were obtained for the different parts of the model M_k in section 1 are still valid.

2.2 The Variable Metric Bundle Algorithm

The variable metric bundle algorithm can now be stated as a variation of algorithm 1.1.

Algorithm 2.1: Nonconvex Variable Metric Bundle Method with Inexact Information

Select parameters $m \in (0,1), \gamma > 0, q > 0, 0 < t_{min} < \frac{1}{q}$ and a stopping tolerance $\texttt{tol} \geq 0$. Choose a starting point $x^1 \in \mathbb{R}^n$ and compute f_1 and g^1. Set the initial metric matrix $Q_1 = \mathbb{I}$, the initial index set $J_1 := \{1\}$ and the initial prox-center to $\hat{x}^1 := x^1$. Set $\hat{f}_1 = f_1$ and select $t_1 > 0$.
For $k = 1, 2, 3, \ldots$

1. Calculate

$$d^k = \arg\min_{d \in \mathbb{R}^n} \left\{ M_k(\hat{x}^k + d) + \mathtt{i}_X(\hat{x}^k + d) + \frac{1}{2}\left\langle d, \left(Q_k + \frac{1}{t_k}\mathbb{I}\right) d \right\rangle \right\}.$$

2. Set

$$G^k = \sum_{j \in J_k} \alpha_j^k s_j^k,$$
$$C_k = \sum_{j \in J_k} \alpha_j^k c_j^k \text{ and}$$
$$\delta_k = C_k + \left\langle d^k, \left(Q_k + \frac{1}{t_k}\mathbb{I}\right) d^k \right\rangle.$$

If $\delta_k \leq \texttt{tol} \rightarrow$ STOP.

3. Set $x^{k+1} = \hat{x}^k + d^k$.

4. Compute f^{k+1}, g^{k+1}.

If $f^{k+1} \leq \hat{f}^k - m\delta_k \quad \rightarrow$ serious step:

Set $\hat{x}^{k+1} = x^{k+1}, \hat{f}^{k+1} = f^{k+1}$ and select $t_{k+1} > t_{min}$.

Calculate a symmetric matrix Q_{k+1} with $-q\mathbb{I} \prec Q_{k+1} \prec q\mathbb{I}$ and adjust t_{k+1} such

$Q_{k+1} + \frac{1}{t_{k+1}} \mathbb{I}$ that $Q_{k+1} + \frac{1}{t_{k+1}}\mathbb{I} \succ 0$.

Otherwise $\quad \rightarrow$ nullstep:

$\hat{f}^{k+1} = \hat{f}^k$ Set $\hat{x}^{k+1} = \hat{x}^k, \hat{f}^{k+1} = f^{k+1}$ and choose $0 < t_{k+1} \leq t_k$.

5. Select the new bundle index set J_{k+1}. Calculate

$$\eta_{k+1} = \max\left\{ \max_{j \in J_{k+1}, x^j \neq \hat{x}^{k+1}} \frac{-2e_j^{k+1}}{|x^j - \hat{x}^{k+1}|^2}, 0 \right\} + \gamma$$

and c_j^{k+1} for all $j \in J^{k+1}$. Update the model M^{k+1}.

2.3 Convergence Analysis

In this section the convergence properties of the new method are analyzed. We do this the same way it is done by Hare et al. in [4].

In the paper all convergence properties are first stated in [4, Lemma 5]. It is then shown that all sequences generated by the method meet the requirements of this lemma which we repeat here for convenience.

Lemma 2.1 ([4, Lemma 5]) *Suppose that the cardinality of the set $\{j \in J^k \mid \alpha_j^k > 0\}$ is uniformly bounded in k.*

(i) If $C^k \to 0$ as $k \to \infty$, then

$$\sum_{j \in J^k} \alpha_j^k \|x^j - \hat{x}^k\| \to 0 \text{ as } k \to \infty.$$

(ii) If additionally for some subset $K \subset \{1, 2, \dots\}$,

$$\hat{x}^k \to \bar{x}, S^k \to \bar{S} \text{ as } K \ni k \to \infty, \text{ with } \{\eta_k \mid k \in K\} \text{ bounded,}$$

then we also have

$$\bar{S} \in \partial f(\bar{x}) + B_{\bar{\theta}}(0).$$

(iii) If in addition $S^k + \nu^k \to 0$ as $K \in k \to \infty$, then $\bar{x}$ satisfies the approximate stationarity condition

$$0 \in (\partial f(\bar{x}) + \partial \mathbf{i}_X(\bar{x})) + B_{\bar{\theta}}(0). \tag{2.6}$$

(iv) Finally if f is also lower-C^1, then for each $\varepsilon > 0$ there exists $\rho > 0$ such that

$$f(y) \geq f(\bar{x}) - (\bar{\theta} + \varepsilon)\|y - \bar{x}\| - 2\bar{\sigma}, \quad \text{for all } y \in X \cap B_\rho(\bar{x}). \tag{2.7}$$

As neither the stabilization nor the descent test are involved in the proof of Lemma 2.1 it is the same as in [4].

We prove now that also the variable metric version of the algorithm fulfills all requirements of Lemma 2.1. The proof is divided into two parts. The first case covers the case of infinitely many serious steps, the second one considers infinitely many null steps after a finite number of serious steps.

For both proofs the equivalence of norms is used between the Euclidean norm and the norm $\|\cdot\|_{Q_k + \frac{1}{t_k}\mathbb{I}}$. As the algorithm assures, that $-q\mathbb{I} \prec Q_k \prec q\mathbb{I}$ for some $q > 0$, the step size can be bounded from below by $t_{min} = \frac{1}{q} - \varepsilon$ for a small $\varepsilon > 0$. This means, the matrix $Q_k + \frac{1}{t_k}\mathbb{I}$ is uniformly bounded for all k. Because the matrix is also symmetric it can be used to define a scalar product which induces indeed the norm $\|\cdot\|_{Q_k + \frac{1}{t_k}\mathbb{I}}$.

The first part of the convergence proof is now stated.

Theorem 2.2 (c.f. [4, Theorem 6, p. 14]) *Let algorithm 2.1 generate an infinite number of serious steps. Then $\delta_k \to 0$ as $k \to \infty$. Let the sequence $\{\eta_k\}$ be bounded. As $k \to \infty$ we have $C_k \to 0$, and for every accumulation point $\bar{x}$ of $\{\hat{x}^k\}$ there exists $\bar{S}$ such that $S^k \to \bar{S}$ and $S^k + \nu^k \to 0$. In particular if the cardinality of $\{j \in J^k \mid \alpha_j^k > 0\}$ is* *uniformly bounded in k then the conclusions of Lemma 2.1 hold.*

The proof is very similar to the one stated in [4] but minor changes have to be made due to the different formulation of the nominal decrease δ_k.

Proof: At each serious step we have

$$\hat{f}_{k+1} \leq \hat{f}_k - m\delta_k, \tag{2.8}$$

where $m,\ \delta_k > 0$. From this follows that the sequence $\{\hat{f}_k\}$ is nonincreasing. Since the sequence $\{\hat{x}^k\}$ lies in the compact set X and f is continuous the sequence $\{f(\hat{x}^k)\}$ is bounded. With $|\sigma_k| < \bar{\sigma}$ also the sequence $\{f(\hat{x}^k) + \sigma_k\} = \{\hat{f}_k\}$ is bounded. Considering also the fact that $\{\hat{f}_k\}$ is nonincreasing one can conclude that it converges.
From (2.8) follows that

2.8 gilt nur für serious steps. Das ist auch im Paper von Hare et.al falsch. Man muss hier nur die Folge der serious steps betrachten.

$$0 \leq m \sum_{k=1}^{l} \delta_k \leq \sum_{k=1}^{l} \left(\hat{f}_k - \hat{f}_{k+1}\right),$$

so letting $l \to \infty$,

$$0 \leq m \sum_{k=1}^{\infty} \delta_k \leq \hat{f}_1 - \underbrace{\lim_{k\to\infty} \hat{f}_k}_{\neq \pm\infty}.$$

This yields

$$\sum_{k=1}^{\infty} \delta_k = \sum_{k=1}^{\infty} \left(C^k + \left\langle d^k, \left(Q_k + \frac{1}{t_k}\mathbb{I}\right) d^k \right\rangle\right) < \infty.$$

Hence, $\delta_k \to 0$ as $k \to \infty$. All quantities above are nonnegative due to positive definiteness of $Q_k + \frac{1}{t_k}\mathbb{I}$ and $C_k \geq 0$ so it also holds that

$$C_k \to 0 \quad \text{and} \quad \left\langle d^k, \left(Q_k + \frac{1}{t_k}\mathbb{I}\right) d^k \right\rangle \to 0.$$

Finally we need to show that for any accumulation point $\bar{x}$ of the sequence $\{\hat{x}^k\}$ holds $S^k \to \bar{S}$ and $S^k + \nu^k \to 0$ for $K \ni k \to \infty$ and the suitable subsequence $K \subset \{1, 2, \ldots\}$. Let $\hat{K}' \subset \{1, 2, \ldots\}$ denote the subset such that the sequence $\{\hat{x}^{k'}\}$ converges to its accumulation point $\bar{x}$ for $k' \in K'$. Let then $\hat{K}' \subset K'$ be the subset of iterates where serious steps were done. From $\{\hat{x}^{k'}\}_{k' \in K'} \to \bar{x}$ follows that $d^{\hat{k}'} = \hat{x}^{\hat{k}'+1} - \hat{x}^{\hat{k}'} \to 0$ for $\hat{K}' \ni \hat{k}' \to \infty$. The step size t_k is bounded below by $t_{min} > 0$ so $\liminf_{k\to\infty} t_k > 0$.

Because of this and because the eigenvalues of Q_k are bounded the expression

$$S^{\hat{k}'} + \nu^{\hat{k}'} = \left(Q_{\hat{k}'} + \frac{1}{t_{\hat{k}'}}\mathbb{I}\right) d^{\hat{k}'} \to 0 \quad \text{for} \quad \hat{K}' \ni \hat{k}' \to \infty$$

because

$$\|S^{\hat{k}'} + \nu^{\hat{k}'}\| = \left\| \left(Q_{\hat{k}'} + \frac{1}{t_{\hat{k}'}}\mathbb{I}\right) d^{\hat{k}'} \right\| \leq \underbrace{\left(\|Q_{\hat{k}'}\| + \frac{1}{t_{\hat{k}'}}\right)}_{\text{bounded}} \underbrace{\|d^{\hat{k}'}\|}_{\to 0}. \; (*)$$

? Wie genau ist $\|Q_{\hat{k}'}\|$ beschränkt für alle $\hat{k}'$?

Es g[…] $\|Q_{\hat{k}'}\| \leq q_1$ aber das musst du bewei[…] Dann kannst du einfach schreiben

$(*) \leq \left(q_1 + \frac{1}{t_{min}}\right) \|d^{\hat{k}'}\|$

The implication $S^k \to \bar{S}$ for $k \in K$ follows from local Lipschitz continuity of f. By Rademacher's theorem this property yields that on any open set U the function f is differentiable except on a set U_{nd} of zero Lebesgue measure. As f is also Lipschitz continuous on any closed set containing such an open set U, the gradient of f on U is bounded. Let $X \subset U$. By theorem 2.5.1 in [1, p. 63] the subdifferential of f at any point $x^k \in U$ is the convex hull of the limits of gradients ∇f of f on the set $U \setminus U_{\text{nd}}$

$$\partial f(x^k) = \text{conv}\{\lim \nabla f(y) \mid y \to x^k, y \notin U_{\text{nd}}\}.$$

This means that also all subgradients on U are bounded. As the subgradient error is assumed to be bounded by $\bar{\theta}$ the set of approximate subgradients $\{g^j, j \in J^k\}$ contained in the bundle is bounded as well. From this follows that also the augmented subgradients $s_j^k = g^j + \eta_k(x^j - \hat{x}^k)$ are bounded because η^k is bounded by assumption and $x^j, \hat{x}^k \in X$. Defining $s := \max_{k,j} \|s_j^k\|$ this yields that

$$\|S^k\| = \left\| \sum_{j \in J^k} \alpha_j^k s_j^k \right\| \leq \sum_{j \in J^k} \|\alpha_j^k\| \underbrace{\|s_j^k\|}_{\leq s \in \mathbf{R}} \leq s \underbrace{\sum_{j \in J^k} \alpha_j^k}_{=1} < \infty \quad \forall k.$$

It follows that the sequence S^k is bounded. By the Bolzano-Weierstrass theorem [7, p. 51] every bounded sequence has a convergent subsequence. Let $K \subset \hat{K}'$ denote the index set of this converging subsequence and $\bar{S}$ the corresponding accumulation point. Then finally $S^k \to \bar{S}$ for $K \ni k \to \infty$.

$\square$

For the case that only finitely many serious steps are executed we need the following result:

Whenever x^{k+1} is as declared a null step, a simple calculation shows that the relation

$$-c_{k+1}^{k+1} + \left\langle s_{k+1}^{k+1}, x^{k+1} - \hat{x}^k \right\rangle = f_{k+1} - \hat{f}_k + \underbrace{\frac{\eta_{k+1}}{2}\|x^{k+1} - \hat{x}^k\|^2}_{\geq 0} > -m\delta_k \tag{2.9}$$

holds. The exact derivation of (2.9) is also given in [4, p. 16].

Another relation that is used a few times throughout the proof is the estimate

$$\left\langle \nu^k, d^k \right\rangle \geq 0. \tag{2.10}$$

It follows from the subgradient inequality for the convex function $\mathtt{i}_X$ at the point x^{k+1}. As $\nu^k \in \partial \mathtt{i}_X(x^{k+1})$ it holds $\mathtt{i}_X(y) - \mathtt{i}_X(x^{k+1}) \geq \left\langle \nu^k, y - x^{k+1} \right\rangle$ for all $y \in X$ and as $d^k = x^{k+1} - \hat{x}^k$ and $x^{k+1}, \hat{x}^k \in X$ it follows

$$0 = \underbrace{\mathtt{i}_X(\hat{x}^k)}_{=0} - \underbrace{\mathtt{i}_X(x^{k+1})}_{=0} \geq \left\langle \nu^k, \hat{x}^k - x^{k+1} \right\rangle = -\left\langle \nu^k, d^k \right\rangle$$

yielding inequality (2.10) above.

Theorem 2.3 (c.f. [4, Theorem 7, p. 16]) *Let a finite number of serious iterates be followed by infinite null steps. Let the sequence $\{\eta_k\}$ be bounded. Then $\{x^k\} \to \hat{x}$, $\delta_k \to 0$, $C_k \to 0$, $S^k + \nu^k \to 0$ and there exist $K \subset \{1, 2, ...\}$ and $\bar{S}$ such that $S^k \to \bar{S}$ as $K \ni k \to \infty$.*
In particular if the cardinality of $\{j \in J^k \mid \alpha_j^k > 0\}$ is uniformly bounded in k then the conclusions of Lemma 2.1 hold for $\bar{x} = \hat{x}$.

Proof: Let k be large enough such that $k \geq \bar{k}$, where $\bar{k}$ is the iterate of the last serious step. Let $\hat{x} := \hat{x}^{\bar{k}}$ and $\hat{f} := \hat{f}_{\bar{k}}$ be fixed. The matrix Q_k is also fixed and denoted as $Q := Q_{\bar{k}}$. Define the optimal value of the k'th subproblem (2.3) for $k > \bar{k}$ by

$$\Psi_k := M_k(x^{k+1}) + \frac{1}{2}\left\langle d^k, \left(Q + \frac{1}{t_k}\mathbb{I}\right) d^k \right\rangle. \tag{2.11}$$

It is first shown that the sequence $\{\Psi_k\}$ is bounded above. From definition (1.6) and relation (1.10) follows

$$A_k(\hat{x}) = M_k(x^{k+1}) - \left\langle S^k, d^k \right\rangle \leq M_k(\hat{x}).$$

Using (2.4) for the third equality and (2.10) in the first inequality one obtains

$$
\begin{aligned}
\Psi_k + \frac{1}{2}\left\langle d^k, \left(Q + \frac{1}{t_k}\mathbb{I}\right) d^k \right\rangle &= A_k(\hat{x}) + \left\langle S^k, d^k \right\rangle + \left\langle d^k, \left(Q + \frac{1}{t_k}\mathbb{I}\right) d^k \right\rangle \\
&= A_k(\hat{x}) + \left\langle S^k + \langle Q + \frac{1}{t_k}\mathbb{I}, d^k \rangle, d^k \right\rangle \\
&= A_k(\hat{x}) - \left\langle \nu^k, d^k \right\rangle \\
&\leq A_k(\hat{x}) \\
&\leq M_k(\hat{x}) \\
&= \hat{f}.
\end{aligned}
$$

By boundedness of d^k and boundedness and positive definiteness $Q + \frac{1}{t_k}\mathbb{I}$ this yields that $\Psi_k \leq \Psi_k + \frac{1}{2}\|d^k\|^2_{Q+\frac{1}{t_k}\mathbb{I}} \leq \hat{f}$, so the sequence $\{\Psi_k\}$ is bounded above. In the next step it is shown that $\{\Psi_k\}$ is increasing. By noting that $x^{k+2} = \hat{x} + d^{k+1}$, as the proximal center does not change in the null step case, we obtain

$$
\begin{aligned}
\Psi_{k+1} &= M_{k+1}(x^{k+2}) + \frac{1}{2}\left\langle d^{k+1}, \left(Q + \frac{1}{t_{k+1}})\mathbb{I}\right) d^{k+1} \right\rangle \\
&\geq A_k(\hat{x} + d^{k+1}) + \frac{1}{2}\left\langle d^{k+1}, \left(Q + \frac{1}{t_k}\mathbb{I}\right) d^{k+1} \right\rangle \\
&= M_k(x^{k+1}) + \left\langle S^k, d^{k+1} - d^k \right\rangle + \frac{1}{2}\left\langle d^{k+1}, \left(Q + \frac{1}{t_k}\mathbb{I}\right) d^{k+1} \right\rangle \\
&= M_k(x^{k+1}) + \left\langle -\left(Q + \frac{1}{t_k}\mathbb{I}\right) d^k - \nu^k, d^{k+1} - d^k \right\rangle + \frac{1}{2}\left\langle d^{k+1}, \left(Q + \frac{1}{t_k}\mathbb{I}\right) d^{k+1} \right\rangle \\
&= \Psi_k - \frac{1}{2}\left\langle d^k, \left(Q + \frac{1}{t_k}\mathbb{I}\right) d^k \right\rangle + \frac{1}{2}\left\langle d^{k+1}, \left(Q + \frac{1}{t_k}\mathbb{I}\right) d^{k+1} \right\rangle \\
&\qquad - \left\langle d^k, \left(Q + \frac{1}{t_k}\mathbb{I}\right)\left(d^{k+1} - d^k\right) \right\rangle - \left\langle \nu^k, d^{k+1} - d^k \right\rangle \\
&= \Psi_k + \frac{1}{2}\left\langle d^k, \left(Q + \frac{1}{t_k}\mathbb{I}\right) d^k \right\rangle + \frac{1}{2}\left\langle d^{k+1}, \left(Q + \frac{1}{t_k}\mathbb{I}\right) d^{k+1} \right\rangle \\
&\qquad - \left\langle d^k, \left(Q + \frac{1}{t_k}\mathbb{I}\right) d^{k+1} \right\rangle - \left\langle \nu^k, x^{k+2} - x^{k+1} \right\rangle \\
&\geq \Psi_k + \frac{1}{2}\left\langle \left(d^{k+1} - d^k\right), \left(Q + \frac{1}{t_k}\mathbb{I}\right)\left(d^{k+1} - d^k\right) \right\rangle \\
&= \Psi_k + \frac{1}{2}\underbrace{\|d^{k+1} - d^k\|^2_{Q+\frac{1}{t_k}\mathbb{I}}}_{\geq 0}.
\end{aligned}
\tag{2.12}
$$

Here the first inequality comes from (1.10) and the fact that $t_{k+1} \le t_k$ for null steps. The second equality follows from (1.6), the fourth equality by (2.4) and (2.11) and the last inequality holds by the subgradient inequality for $\nu^k \in \mathtt{i}_X(x^{k+1})$ and the fact that $x^{k+1}, x^{k+2} \in X$.

As Q is fixed in null steps and $\liminf_{k\to\infty} t_k > 0$ (because $t_k > t_{min} > 0$ for all k) the sequence $\{\Psi_k\}$ is increasing and bounded from above. It is therefore convergent. ?

Looking again at (2.12) and taking into account that $1/t_k \ge 1/t_{\bar{k}}$ in the null step case we have

$$
\begin{aligned}
\Psi_{k+1} - \Psi_k &\ge \frac{1}{2}\|d^{k+1} - d^k\|^2_{Q+\frac{1}{t_k}\mathbb{I}} \\
&\ge \frac{1}{2}\|d^{k+1} - d^k\|^2_{Q+\frac{1}{t_{\bar{k}}}\mathbb{I}}. > 0
\end{aligned}
$$

$\Rightarrow \{\Psi_k\}$ increasing

$\Rightarrow$ -"- convergent

As the sequence $\{\Psi_k\}$ is converging this yields

$$
|\Psi_{k+1} - \Psi_k| \to 0 \quad \Rightarrow \quad \|d^{k+1} - d^k\| \to 0 \text{ for } k \to \infty \tag{2.13}
$$

due to the equivalence of norms.

By the last line in (2.5) and the fact that $\hat{f} = M_k(\hat{x})$ for all $k > \bar{k}$ we have

$$
\begin{aligned}
\hat{f} &= M_k(\hat{x}) + \delta_k - C_k - \left\langle d^k, \left(Q + \frac{1}{t_k}\mathbb{I}\right) d^k \right\rangle \\
&= M_k(\hat{x}) - \hat{f} + M_k(x^{k+1}) + \delta_k - \left\langle S^k, d^k \right\rangle - \left\langle d^k, \left(Q + \frac{1}{t_k}\mathbb{I}\right) d^k \right\rangle \\
&= \delta_k + M_k(\hat{x} + d^k) + \left\langle \nu^k, d^k \right\rangle \\
&\ge \delta_k + M_k(\hat{x} + d^k),
\end{aligned}
$$

where the second equality is by (1.9), the third holds because of relation (2.4) and the last inequality is given by (2.10). Therefore

$$
\delta^{k+1} \le \hat{f} - M_{k+1}(\hat{x} + d^{k+1}). \tag{2.14}
$$

By assumption (1.11) on the model, written for $d = d^{k+1}$,

$$-\hat{f}_{k+1} + c_{k+1}^{k+1} - \left\langle s_{k+1}^{k+1}, d^{k+1} \right\rangle \geq -M_{k+1}(\hat{x} + d^{k+1}).$$

In the null step case it holds $\hat{f}_{k+1} = \hat{f}$ so combining condition (2.9) and the inequality above, one obtains that

$$m\delta_k + \left\langle s_{k+1}^{k+1}, d^k - d^{k+1} \right\rangle \geq \hat{f} - M_{k+1}(\hat{x} + d^{k+1}).$$

In combination with (2.14) this yields

$$0 \leq \delta_{k+1} \leq m\delta_k + \left\langle s_{k+1}^{k+1}, d^k - d^{k+1} \right\rangle \leq m\delta_k + \left| \left\langle s_{k+1}^{k+1}, d^k - d^{k+1} \right\rangle \right|. \tag{2.15}$$

For the next step Lemma 3 and the corollary below it from [15, p. 45] are used. They state that for

$$u_{k+1} \leq q u_k + a_k, \quad q < 1, \quad a_k \geq 0, \quad a_k \to 0 \text{ and } u_k \geq 0$$

it holds $u_k \to 0$.

Taking the first and the last part of inequality (2.15) we can identify $u_k = \delta_k \geq 0$, $q = m \in (0,1)$ and $a_k = \left| \left\langle s_{k+1}^{k+1}, d^k - d^{k+1} \right\rangle \right| \geq 0$. To show that $a_k \to 0$ recall (2.13) and that the augmented subgradient s_{k+1}^{k+1} is bounded due to local Lipschitz continuity of f and boundedness of $\{\eta_k\}$ by the same argumentation as in the case of infinitely many serious steps.

The lemma then gives that

$$\lim_{k\to\infty} \delta_k = \lim_{k\to\infty} C_k + \left\langle d^k, \left(Q_k + \frac{1}{t_k}\mathbb{I}\right) d^k \right\rangle = 0.$$

By the fact that $C_k \geq 0$ for all k and positive definiteness of $Q_k + \frac{1}{t_k}\mathbb{I}$ it follows that all summands above are nonnegative and hence $C_k \to 0$ as $k \to \infty$. Since $Q + \frac{1}{t_k}\mathbb{I}$ is also bounded due to $\liminf_{k\to\infty} t_k > 0$ and the bounded eigenvalues of Q we have

As $Q_k = Q$ and $t_k \leq t_{\bar{k}}$ for k sufficiently large, $\langle d^k, (Q_k + \frac{1}{t_k}) d^k \rangle \geq \langle d^k, (Q + \frac{1}{t_{\bar{k}}}) d^k \rangle \geq c\|d^k\|^2 \to 0$

$$\|d^k\|^2_{Q+\frac{1}{t_k}\mathbb{I}} \geq c\|d^k\|^2 \to 0$$

for a constant $c \in \mathbb{R}$ by the equivalence of norms.

This means that $d^k \to 0$ for $k \to \infty$ and therefore $\lim_{k\to\infty} x^k = \hat{x}$. It also follows that

$\|S^k + \nu^k\| \to 0$ as $k \to \infty$ because of

$$\|S^k + \nu^k\| = \left\| \left(Q_k + \frac{1}{t_k}\mathbb{I} \right) d^k \right\| \leq \underbrace{\left(\|Q_k\| + \frac{1}{t_k} \right)}_{\text{bounded}} \underbrace{\|d^k\|}_{\to 0} \to 0.$$

Wieso ist das beschränkt?

Wenn nur nullsteps erfolge. kann $t_k \to 0$ gehen.

By the same arguments as in the proof of theorem 2.2 the local Lipschitz property of the objective function f and boundedness of the subgradient errors θ_k yield boundedness of the sequence S^k. Passing to some subsequence $K \subset \{1, 2, \ldots\}$ if necessary we can therefore conclude that the sequence $\{S^k\}_{k \in K}$ converges to some $\bar{S}$ and as $\hat{x}^k = \bar{x}$ for all $k \geq \bar{k}$ all requirements of Lemma 2.1 are fulfilled.

□

Remark: In case the matrix Q_k is also updated in null steps the proof still holds as long as the assumptions on boundedness of Q_k and positive definiteness of $Q_k + \frac{1}{t_k}\mathbb{I}$ are still valid.

Remark: All results deduced in section 1.2 are still valid for this algorithm as they do not depend on the kind of stabilization used.

2.4 Updating the Metric

In [14] and [13] it is not specified how the matrices Q_k are chosen. For convergence it is necessary that Q_k is symmetric and its eigenvalues are bounded. Here we present some possibilities to update the metric matrix Q_k such that it fulfills both conditions.

Most of the presented updates are based on the BFGS-update formula (named after Broyden, Goldfarb, Fletcher and Shanno)

$$\tilde{Q}_{k+1} = Q_k + \frac{y^k {y^k}^\top}{\langle y^k, d^k \rangle} - \frac{Q_k d^k (Q_k d^k)^\top}{\langle d^k, Q_k d^k \rangle}. \tag{2.16}$$

Usually y^k is defined as the difference of the last two gradients of f. To adapt the formula to the nondifferentiable case the difference $y^k := g^{k+1} - g^k$ of two (approximate) subgradients of f is taken instead as proposed in [2]. The starting matrix is $Q_1 = \mathbb{I}$.

By definition the BFGS update is symmetric. To assure boundedness of the matrix Q_{k+1} the updates can be manipulated in the following ways: