Mathematical Programming 31 (1985) 269-285
North-Holland

ON THE SUPERLINEAR CONVERGENCE OF A TRUST REGION ALGORITHM FOR NONSMOOTH OPTIMIZATION

Y. YUAN

Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge CB3 9EW, England

Received 3 November 1983
Revised manuscript received 15 September 1984

It is proved that the second order correction trust region algorithm of Fletcher [5] ensures superlinear convergence if some mild conditions are satisfied.

Key words: Trust Region Algorithms, Nonsmooth Optimization, Superlinear Convergence.

1. Introduction

Fletcher [5] presents a trust region algorithm with a second order correction to solve the following composite optimization problem:

$$\min_{x} \phi(x) \equiv f(x) + h(c(x)), \tag{1.1}$$

where $f(x)$ from $\mathbb{R}^n$ to $\mathbb{R}$ and $c(x)$ from $\mathbb{R}^n$ to $\mathbb{R}^m$ are twice continuously differentiable functions and $h(c)$ from $\mathbb{R}^m$ to $\mathbb{R}$ is a polyhedral convex function of the form

$$h(c) = \max_{1 \leq i \leq l} (h_i^{\mathrm{T}} c + b_i), \tag{1.2}$$

h_i and b_i being given vectors and constants respectively. The algorithm is based on a model algorithm of Fletcher [3] and it is iterative. At the begining of the k-th iteration, $x^{(k)}$, $\lambda^{(k)}$ and ρ^k are available, where $x^{(k)}$ is an estimate of the solution of (1.1), $\lambda^{(k)}$ is an estimate of the Lagrangian multipliers at the solution and $\rho^{(k)}$ is a trust region bound. The algorithm requires the solution of the following subproblem:

$$\min \psi^{(k)}(d) \equiv q^{(k)}(d) + h(c(x^{(k)}) + A(x^{(k)})d) \tag{1.3}$$

subject to

$$\|d\| \leq \rho^{(k)}, \tag{1.4}$$

where

$$q^{(k)}(d) = f(x^{(k)}) + \nabla^{\mathrm{T}} f(x^{(k)}) d + \tfrac{1}{2} d^{\mathrm{T}} W^{(k)} d,$$

$$W^{(k)} = \nabla^2 f(x^{(k)}) + \sum_{i=1}^{m} \lambda_i^{(k)} \nabla^2 c_i(x^{(k)}), \tag{1.5}$$

$\|\cdot\|$ is any given norm and $A = \nabla^T c \in \mathbb{R}^{m\times n}$ is the Jacobian of c. Let $d^{(k)}$ be a solution of (1.3) and (1.4); then

$$r^{(k)} = \frac{\phi(x^{(k)}) - \phi(x^{(k)} + d^{(k)})}{\psi^{(k)}(\mathbf{0}) - \psi^{(k)}(d^{(k)})} \tag{1.6}$$

is calculated, which is the ratio between the actual reduction and the predicted reduction of the objective function. On some iterations the algorithm also solves the following 'second order correction' subproblem:

$$\min \hat{\psi}^{(k)}(d) = q^{(k)}(d^{(k)} + d) + h(c(x^{(k)} + d^{(k)}) + A(x^{(k)})d) \tag{1.7}$$

subject to

$$\|d^{(k)} + d\| \leq \rho^{(k)}. \tag{1.8}$$

Let $\hat{d}^{(k)}$ be a solution of (1.7) and (1.8) and define

$$\begin{aligned} r_e^{(k)} &= r^{(k)} + \frac{\hat{\psi}^{(k)}(\mathbf{0}) - \hat{\psi}^{(k)}(\hat{d}^{(k)})}{\psi^{(k)}(\mathbf{0}) - \psi^{(k)}(d^{(k)})}, \\ \hat{r}^{(k)} &= \frac{\phi(x^{(k)}) - \phi(x^{(k)} + d^{(k)} + \hat{d}^{(k)})}{\psi^{(k)}(\mathbf{0}) - \psi^{(k)}(d^{(k)})}. \end{aligned} \tag{1.9}$$

Following Fletcher's notation, we let $\hat{x}^{(k)} = x^{(k)} + d^{(k)}$, $\tilde{x}^{(k)} = \hat{x}^{(k)} + \hat{d}^{(k)}$, $\Delta\psi^{(k)} = \psi^{(k)}(\mathbf{0}) - \psi^{(k)}(d^{(k)})$ and $\Delta\hat{\psi}^{(k)} = \hat{\psi}^{(k)}(\mathbf{0}) - \hat{\psi}^{(k)}(\hat{d}^{(k)})$, and we let g denote ∇f, ∂h denote the subdifferential of h, and $g^{(k)}$ denote $g(x^{(k)})$ etc.

Using the above notation, the details of an iteration of the algorithm are as follows, where k is the iteration number.

Fletcher's Algorithm

Step 1. Evaluate $f^{(k)}$, $c^{(k)}$, $g^{(k)}$, $A^{(k)}$ and $W^{(k)}$;
Solve (1.3)—(1.4) giving $d^{(k)}$;
Evaluate $r^{(k)}$, if $r^{(k)} > 0.75$ go to Step 7.

Step 2. Solve (1.7)—(1.8) giving $\hat{d}^{(k)}$;
Evaluate $r_e^{(k)}$; if $r^{(k)} < 0.25$ go to Step 4.

Step 3. If $r_e^{(k)} \in [0.9, 1.1]$, set $\rho^{(k+1)} = 2\rho^{(k)}$ and go to Step 9, otherwise go to Step 8.

Step 4. If $r_e^{(k)} \notin [0.75, 1.25]$ go to Step 5;
Evaluate $\hat{r}^{(k)}$;
In computation assign $d^{(k)} := d^{(k)} + \hat{d}^{(k)}$, $r^{(k)} := \hat{r}^{(k)}$;
If $r^{(k)} > 0.75$ go to Step 7.
If $r^{(k)} \geq 0.25$ go to Step 8.

Step 5. Set $\rho^{(k+1)} = \alpha_k \|d^{(k)}\|$, $\alpha_k \in [0.1, 0.5]$,
If $r^{(k)} > 0$ go to Step 9.

Step 6. $x^{(k+1)} := x^{(k)}$, generate $\boldsymbol{\lambda}^{(k+1)}$; end of k-th iteration.

Step 7. If $\|d^{(k)}\| < \rho^{(k)}$ go to Step 8;
If $r^{(k)} > 0.9$ then $\rho^{(k+1)} := 4\rho^{(k)}$ else $\rho^{(k+1)} := 2\rho^{(k)}$;
go to Step 9.

Step 8. $\rho^{(k+1)} := \rho^{(k)}$.

Step 9. Set $\boldsymbol{x}^{(k+1)} := \boldsymbol{x}^{(k)} + \boldsymbol{d}^{(k)}$, generate $\boldsymbol{\lambda}^{(k+1)}$; end of k-th iteration.

The value of α_k in Step 5 is an estimate of the number α_k^* which satisfies $\phi(\boldsymbol{x}^{(k)} + \alpha_k^* \boldsymbol{d}^{(k)}) = \min_{0.1 \le \alpha \le 0.5} \phi(\boldsymbol{x}^{(k)} + \alpha \boldsymbol{d}^{(k)})$. Fletcher [5] gives a specific choice of α_k, which does not require any line searches but only depends on the value of $r^{(k)}$. Further, he lets $\boldsymbol{\lambda}^{(k+1)}$ be $\boldsymbol{\lambda}^{(k)}$ in Step 6 and $\boldsymbol{\lambda}^{(k+1)}$ be the multipliers from either the subproblem (1.3)–(1.4) or the subproblem (1.7)–(1.8) in Step 9.

Fletcher [5] shows that if $\{x_k\}$ $(k = 1, 2, \ldots)$ are all in a bounded set in $\mathbb{R}^n$ then $\{x_k\}$ is not bounded away from the Stationary Points, where a stationary point x^* means a point at which

$$f(\boldsymbol{x}^*) + h(\boldsymbol{c}(\boldsymbol{x}^*)) \le f(\boldsymbol{x}^*) + \nabla^{\mathrm{T}} f(\boldsymbol{x}^*)\boldsymbol{d} + h(\boldsymbol{c}(\boldsymbol{x}^*) + \nabla^{\mathrm{T}} \boldsymbol{c}(\boldsymbol{x}^*)\boldsymbol{d})$$

holds for all $\boldsymbol{d} \in \mathbb{R}^n$. The condition that $\{x_k\}$ is bounded is usually satisfied, specifically if x_1 is so chosen that

$$\boldsymbol{x}: \quad \phi(\boldsymbol{x}) < \phi(\boldsymbol{x}_1)$$

is a bounded set. Let x^* be an accumulation point of $\{x_k\}$. Since our interest is in the rate of convergence, we asume that strict complementarity and second order sufficiency conditions are satisfied at $\boldsymbol{x}^*$, and that $\mathrm{Rank}(A^*) = m$. Due to the second order sufficiency condition it is easy to show that $x_k \to x^*$, since x^* is an accumulation point of $\{x_k\}$. We assume that the functions $f(\cdot)$ and $\boldsymbol{c}(\cdot)$ are three times continuously differentiable. Without loss of generality, we also assume that $h(\boldsymbol{c}^*) = \boldsymbol{h}_i^{\mathrm{T}} \boldsymbol{c}^* + b_i$ for all $i = 1, 2, \ldots, I$. Let $\boldsymbol{\lambda}^* \in \partial h^*$ (∂h^* is the convex hull of the vectors $\boldsymbol{h}_i$, $i = 1, 2, \ldots, I$) be the Lagrangian multipliers at $\boldsymbol{x}^*$ such that

$$\boldsymbol{g}^* + (A^*)^{\mathrm{T}} \boldsymbol{\lambda}^* = 0. \tag{1.10}$$

The full rank of A^* ensures the uniqueness of $\boldsymbol{\lambda}^*$ and the second order sufficiency condition states that

$$\boldsymbol{d}^{\mathrm{T}} W^* \boldsymbol{d} > 0 \quad \text{for all } \boldsymbol{d} \in G^* \tag{1.11}$$

where

$$\begin{aligned} G^* &= \left\{ \boldsymbol{d}: \max_{\boldsymbol{\lambda} \in \partial h^*} \boldsymbol{d}^{\mathrm{T}}(\boldsymbol{g}^* + (A^*)^{\mathrm{T}} \boldsymbol{\lambda}) = 0,\ \boldsymbol{d} \ne \boldsymbol{0} \right\}, \\ W^* &= \nabla^2 f(\boldsymbol{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla^2 c_i(\boldsymbol{x}^*). \end{aligned} \tag{1.12}$$

Under the above assumptions, Fletcher [5] proves that, if the trust region bound is inactive for all large k, then the algorithm converges quadratically. Yuan [16], however, gives examples of trust region methods that converge only linearly, so it is important to investigate the effect of the trust region bound when k is large.

Since our result is stongly dependent on the condition $\boldsymbol{\lambda}^{(k)} \to \boldsymbol{\lambda}^*$, and since we are not able to prove that the original choice of $\boldsymbol{\lambda}^{(k)}$ in Fletcher [5] ensures this condition, throughout this paper we assume that the generation of $\boldsymbol{\lambda}^{(k+1)}$ gives the limit:

$$\lambda^{(k)} \to \lambda^*. \tag{1.13}$$

Many suitable methods for estimating Lagrangian multipliers are known, for example Murray and Overton [10, 11]. The condition (1.13) admits many estimation techniques, such as

$$\lambda^{(k+1)} = \arg\min_{\lambda} \|g^{(k)} - (A^{(k)})^{\mathrm{T}}\lambda\|. \tag{1.14}$$

Some lemmas are stated and the main theorem is proved in the following section. The proofs of the lemmas are given in the Appendix.

2. The result

In this section, we demonstrate superlinear convergence of the algorithm by showing that the trust region bound is eventually inactive. To make the proof of the main theorem straightforward, we give some lemmas without proofs, and only prove the main theorem. The proofs of the lemmas can be found in the Appendix. The following Lemma 2.1 is due to Fletcher [5], Lemma 2.2 is a generalization of Lemma 4.3 of Yuan [15], and Lemma 2.3 is the main result which indicates that the trust region bound is bounded away from zero.

Lemma 2.1 (Fletcher [5]). *There exists a positive constant c_1 and a neighbourhood of x^*, such that for all x in the neighbourhood the inequality*

$$\phi(x) - \phi(x^*) \geq c_1\|x - x^*\|^2 \tag{2.1}$$

holds.

Lemma 2.2. *For any given $\alpha_0 \in (0, 1)$, there exists a neighbourhood of x^* such that, for all x in the neighbourhood,*

$$\phi(x) - \min_{\|d\| \leq \|x - x^*\|} [f(x) + g(x)^{\mathrm{T}}d + h(c(x) + A(x)d)] \geq \alpha_0[\phi(x) - \phi(x^*)] \tag{2.2}$$

holds.

Lemma 2.3. *Let $d^{(k)}$ solve (1.3)-(1.4) and $\hat{d}^{(k)}$ solve (1.7)-(1.8), then*

(1) $\|d^{(k)}\| = O(\|x^{(k)} - x^*\|)$;

(2) $\|\hat{d}^{(k)}\| = O(\|x^{(k)} - x^*\|)$;

(3) $\Delta\psi^{(k)} \geq c_2\|d^{(k)}\|^2$ *for some positive constant* c_2;

(4) $r_e^{(k)} \to 1$ *as* $k \to +\infty$;

(5) *For any subsequence* $\{k_j\}$, *if* $r^{(k_j)} \leqslant \alpha_1 < 1$ *for some constant* α_1, *then*

$$\|\hat{d}^{(k_j)}\| = o(\|d^{(k_j)}\|), \tag{2.4}$$

$$\lim_{j\to+\infty} \hat{r}^{(k_j)} = 1. \tag{2.5}$$

Corollary 2.4. *For all sufficiently large* k, $\rho^{(k+1)} \geqslant \rho^{(k)}$.

Proof. Let α_1 in Lemma 2.3 have the value 0.75. Hence for all sufficiently large k, either $r^{(k)} > 0.75$, or $r_e^{(k)} \in [0.9, 1.1]$ and $\hat{r}^{(k)} \in [0.9, 1.1]$. Therefore by the definition of $\rho^{(k+1)}$, we have $\rho^{(k+1)} \geqslant \rho^{(k)}$. □

Theorem 2.5. *If* $\boldsymbol{\lambda}^{(k)} \to \boldsymbol{\lambda}^*$, *the algorithm converges superlinearly.*

Proof. Corollary 2.4 ensures that $\{\rho^{(k)}\}$ is bounded away from zero, therefore for large k, the trust region bound is inactive. So from (2.8) and (2.9) of Fletcher [5] we have that

$$\frac{\|\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^*\|}{\|\boldsymbol{x}^{(k)} - \boldsymbol{x}^*\|} = O(\max[\|\boldsymbol{x}^{(k)} - \boldsymbol{x}^*\|, \|\boldsymbol{\lambda}^{(k)} - \boldsymbol{\lambda}^*\|]). \tag{2.6}$$

Thus the superlinear convergence result follows. □

Corollary 2.6. *If* $\boldsymbol{\lambda}^{(k)}$ *is chosen to be a value of* $\boldsymbol{\lambda}$ *that minimizes* $\|\boldsymbol{g}^{(k)} - (A^{(k)})^{\mathrm{T}}\boldsymbol{\lambda}\|_2$, *then the algorithm either terminates within a finite number of iterations of converges to* $\boldsymbol{x}^*$ *quadratically.*

Proof. Since A^* has full rank, for large k, $A^{(k)}$ also has full rank. Therefore $\boldsymbol{\lambda}^{(k)} = ((A^{(k)})^{\mathrm{T}})^{+}\boldsymbol{g}^{(k)}$. Because $\boldsymbol{\lambda}^* = ((A^*)^{\mathrm{T}})^{+}\boldsymbol{g}^*$ and Rank $(A^*) = m$, we have that $\|\boldsymbol{\lambda}^{(k)} - \boldsymbol{\lambda}^*\| = O(\|\boldsymbol{x}^{(k)} - \boldsymbol{x}^*\|)$. Thus, the quadratic convergence follows from (2.6). □

3. Discussion

By modifying the proof, one can show that the superlinear convergence result remains valid if $W^{(k)}$ is replaced by any matrix B_k that satisfies

$$B_k \to W^*, \quad k \to \infty. \tag{3.1}$$

Further, the same is true if $W^{(k)}$ is replaced by any matrix B_k that tends to W^* along directions $\boldsymbol{d}^{(k)}$, $\hat{\boldsymbol{d}}^{(k)}$ and $\boldsymbol{x}^{(k)} - \boldsymbol{x}^*$, which means

$$\|(B_k - W^*)z\| / \|z\| \to 0 \tag{3.2}$$

for $z = \boldsymbol{d}^{(k)}$, $\hat{\boldsymbol{d}}^{(k)}$ and $\boldsymbol{x}^{(k)} - \boldsymbol{x}^*$, since to ensure the proofs in the Appendix valid we only need that

$$W^{(k)}z = W^*z + o(\|z\|),$$

holds for $z = d^{(k)}$, $\hat{d}^{(k)}$ and $x^{(k)} - x^*$. Therefore the sequence $\{B_k\}$ can be generated by updating using function values and first order derivatives, which is usually much easier than calculating second order derivatives. The PSB formula (Powell [13]) is recommended, which works well in trust region algorithms for smooth optimization (Powell [14]). In order to satisfy (3.2) the PSB formula has the form

$$B_{k+1} = B_k + \frac{\sigma^{(k)}\delta^{(k)\mathrm{T}} + \delta^{(k)}\sigma^{(k)\mathrm{T}}}{\|\delta^{(k)}\|^2} - \frac{\delta^{(k)}\delta^{(k)\mathrm{T}}(\sigma^{(k)\mathrm{T}}\delta^{(k)})}{\|\delta\|^4}, \tag{3.3}$$

where

$$\sigma^{(k)} = \sum_{i=1}^{m} \lambda_i^{(k)}[\nabla c_i(x^{(k)} + \delta^{(k)}) - \nabla c_i(x^{(k)})] + \nabla f(x^{(k)} + \delta^{(k)}) - \nabla f(x^{(k)})$$
$$- B_k\delta^{(k)}, \quad \delta^{(k)} = d^{(k)} \tag{3.4}$$

and $\{\lambda_i^{(k)}\}$ $(i = 1, 2, \ldots, m)$ are estimates of the Lagrangian multipliers λ_i^*. It is expected that $\{B_k\}$ updated by (3.3)-(3.4) ensures (3.2), but we are not able to prove it.

If the norm $\|\cdot\|$ in (1.4) and (1.8) is the infinity norm or the 1-norm, the subproblems (1.3)-(1.4) and (1.7)-(1.8) can be solved by using techniques such as in Bartels, Conn and Sinclair [1], and they can also be rewritten as linearly constrained quadratic programming calculations. For the solving of *QP* problems, see Fletcher [2], Gill et al. [9], Gill and Murray [8] and Goldfarb and Idnani [6].

Instead of updating the full matrix B_k, as pointed out by a referee, we can update just the projected matrix. For details of the projected update methods, see, for example, Fletcher [4], Gill and Murray [7] and Nocedal and Overton [12].

Acknowledgements

The author is greatly indebted to Professor M.J.D. Powell for his constant guidance and encouragement, and for his studying the early drafts of this paper. Without his help, this paper never would have been possible. The author wishes to thank Dr. R. Fletcher for some helpful discussions and corrections. I am also grateful to the referees for their useful comments and suggestions which lead to improvements.

Appendix – Proofs of Lemma 2.2 and Lemma 2.3

Proof of Lemma 2.2. If the lemma is invalid, then there exist $\alpha_0 \in (0, 1)$, and a sequence $\{x^{(k)}\}$ $(k = 1, 2, \ldots)$ such that

$$x^{(k)} \to x^* \tag{A.1}$$

and

$$\phi(x^{(k)}) - \min_{\|d\| \le \|x^{(k)} - x^*\|} [f(x^{(k)}) + g^{(k)\mathrm{T}} d + h(c^{(k)} + A^{(k)} d)]$$
$$< \alpha_0[\phi(x^{(k)}) - \phi(x^*)]. \tag{A.2}$$

Since

$$\phi(x^{(k)}) - [f(x^{(k)}) + g^{(k)\mathrm{T}}(x^* - x^{(k)}) + h(c^{(k)} + A^{(k)}(x^* - x^{(k)})]$$
$$= \phi(x^{(k)}) - \phi(x^*) + \mathrm{O}(\|x^* - x^{(k)}\|^2), \tag{A.3}$$

we have that, using (A.2)

$$\phi(x^{(k)}) - \phi(x^*) = \mathrm{O}(\|x^{(k)} - x^*\|^2). \tag{A.4}$$

Hence

$$g^{*\mathrm{T}}(x^{(k)} - x^*) + \max_{1 \le i \le l} h_i^{\mathrm{T}} A^*(x^{(k)} - x^*) = \mathrm{O}(\|x^{(k)} - x^*\|^2), \tag{A.5}$$

using the assumption that $h(c^*) = h_i^{\mathrm{T}} c^* + b_i$ for all i, $1 \le i \le l$. Moreover, the strict complementarity and the definition of $h(\cdot)$ imply that there exist $u_i^* > 0$ such that

$$g^{*\mathrm{T}} + \sum_{i=1}^{l} u_i^* h_i^{\mathrm{T}} A^* = \mathbf{0}, \tag{A.6}$$

and

$$\sum_{i=1}^{l} u_i^* = 1. \tag{A.7}$$

From (A.5)—(A.7), it follows that

$$\liminf_{k \to \infty} \frac{g^{*\mathrm{T}}(x^{(k)} - x^*) + \min_{1 \le i \le l} h_i^{\mathrm{T}} A^*(x^{(k)} - x^*)}{\|x^{(k)} - x^*\|^2} > -\infty, \tag{A.8}$$

which gives

$$g^{*\mathrm{T}}(x^{(k)} - x^*) + h_i^{\mathrm{T}} A^*(x^{(k)} - x^*) = \mathrm{O}(\|x^{(k)} - x^*\|^2). \tag{A.9}$$

for all $i = 1, 2, \ldots, l$.

Consider

$$\min_{\|d\| \le \|x^{(k)} - x^*\|} [f(x^{(k)}) + g^{(k)\mathrm{T}} d + h(c^{(k)} + A^{(k)} d)]$$
$$= \min_{\|d\| \le \|x^{(k)} - x^*\|} \Big\{ \max_{1 \le i \le l} [f(x^*) + h_i^{\mathrm{T}} c(x^*) + b_i + \tfrac{1}{2} g^{*\mathrm{T}}(x^{(k)} - x^*)$$
$$+ \tfrac{1}{2} g^{(k)\mathrm{T}}(x^{(k)} - x^*) + g^{(k)\mathrm{T}} d + h_i^{\mathrm{T}}(\tfrac{1}{2} A^*(x^{(k)} - x^*)$$
$$+ \tfrac{1}{2} A^{(k)}(x^{(k)} - x^*) + A^{(k)} d)] \Big\} + \mathrm{o}(\|x^{(k)} - x^*\|^2)$$
$$\le \min_{\|d\| \le \|x^{(k)} - x^*\|} \Big\{ \max_{1 \le i \le l} [f(x^*) + h_i^{\mathrm{T}} c^* + b_i + \tfrac{1}{2}(g^* + A^{*\mathrm{T}} h_i)^{\mathrm{T}}(x^{(k)} - x^*)$$
$$+ \tfrac{1}{2}(g^{(k)} + (A^{(k)})^{\mathrm{T}} h_i)^{\mathrm{T}} d] \Big\} + \mathrm{o}(\|x^{(k)} - x^*\|^2), \tag{A.10}$$

where the last part is obtained by forcing d in the middle part to have the form $\frac{1}{2}(\tilde{d}+x^*-x^{(k)})$ for $\|\tilde{d}\| \leq \|x^{(k)}-x^*\|$. Let $B(x)$ be the $l \times n$ matrix

$$B(x)=\begin{bmatrix} g^{\mathrm{T}}(x)+h_1^{\mathrm{T}}A(x) \\ g^{\mathrm{T}}(x)+h_2^{\mathrm{T}}A(x) \\ \vdots \\ g^{\mathrm{T}}(x)+h_l^{\mathrm{T}}A(x) \end{bmatrix}. \tag{A.11}$$

It follows from (A.9) that

$$B(x^*)(x^{(k)}-x^*)=\mathrm{O}(\|x^{(k)}-x^*\|^2). \tag{A.12}$$

Hence the vector

$$\dot{d}^{(k)}=-(B(x^*))^+B(x^*)(x^{(k)}-x^*), \tag{A.13}$$

satisfies

$$\|\dot{d}^{(k)}\|=\mathrm{O}(\|x^{(k)}-x^*\|^2), \tag{A.14}$$

and

$$B(x^*)\dot{d}^{(k)}=-B(x^*)(x^{(k)}-x^*). \tag{A.15}$$

Therefore

$$B(x^*)(x^{(k)}-x^*)+B(x^{(k)})\dot{d}^{(k)}=\mathrm{o}(\|x^{(k)}-x^*\|^2). \tag{A.16}$$

Because, for sufficiently large k, (A.14) implies $\|\dot{d}^{(k)}\| \leq \|x^{(k)}-x^*\|$, it follows from (A.10), (A.14) and (A.16) that

$$\begin{aligned}\min_{\|d\|\leq\|x^{(k)}-x^*\|} & [f(x^{(k)}+g^{(k)\mathrm{T}}d+h(c^{(k)}+A^{(k)}d)] \\ & \leq f(x^*)+h(c^*)+\tfrac{1}{2}\|B(x^*)(x^{(k)}-x^*)+B(x^{(k)})\dot{d}^{(k)}\|_\infty+\mathrm{o}(\|x^{(k)}-x^*\|^2) \\ & = f(x^*)+h(c^*)+\mathrm{o}(\|x^{(k)}-x^*\|^2). \end{aligned} \tag{A.17}$$

(A.2) and (A.17) imply that

$$(1-\alpha_0)[\phi(x^{(k)}-\phi(x^*)]\leq \mathrm{o}(\|x^{(k)}-x^*\|^2), \tag{A.18}$$

which contradicts Lemma 2.1. Therefore Lemma 2.2 is true. □

Proof of Lemma 2.3. If (1) is not true, then there exist k_j ($j=1,2,\ldots$) such that

$$\|x^{(k_j)}-x^*\|=\mathrm{o}(\|d^{(k_j)}\|) \tag{A.19}$$

and

$$\psi^{(k_j)}(d^{(k_j)})=\min_{\|d\|\leq\rho^{(k_j)}}\psi^{(k_j)}(d). \tag{A.20}$$

Thus we have, for sufficiently large k_j,

$$\psi^{(k_j)}(d^{(k_j)})\leq\psi^{(k_j)}(x^*-x^{(k_j)})=\phi(x^*)+\mathrm{O}(\|x^{(k_j)}-x^*\|^2). \tag{A.21}$$

Choosing a subsequence if necessary, we assume

$$d^{(k_j)}/\|d^{(k_j)}\| \to d'. \tag{A.22}$$

which gives

$$(x^{(k_j)}+d^{(k_j)}-x^*)/\|d^{(k_j)}\| \to d'. \tag{A.23}$$

Since

$$\phi(x^{(k_j)}+d^{(k_j)}) = \psi^{(k_j)}(d^{(k_j)})+\mathrm{O}(\|d^{(k_j)}\|^2), \tag{A.24}$$

it follows from (A.21), (A.24) and the fact that x^* is a minimum of $\phi(\cdot)$ that

$$\phi(x^{(k_j)}+d^{(k_j)}) = \phi(x^*)+\mathrm{O}(\|x^{(k_j)}+d^{(k_j)}-x^*\|^2). \tag{A.25}$$

Therefore the definition (1.12) of G^*, (A.19), (A.23), (A.25) and the inequality

$$\phi(x)-\phi(x^*) \geq \max_{\lambda\in\partial h^*} (x-x^*)^{\mathrm{T}}(g^*+(A^*)^{\mathrm{T}}\lambda^*)+\mathrm{O}(\|x-x^*\|^2)$$

imply that

$$d' \in G^*, \tag{A.26}$$

so the second order sufficiency condition ensures

$$d'^{\mathrm{T}}W^*d' > 0. \tag{A.27}$$

Hence, using (A.23)

$$(x^{(k_j)}+d^{(k_j)}-x^*)^{\mathrm{T}}W^*(x^{(k_j)}+d^{(k_j)}-x^*) \geq \tfrac{1}{2}d'^{\mathrm{T}}W^*d'\|d^{(k_j)}\|^2 \tag{A.28}$$

for sufficiently large j. Thus, remembering that h is convex and $\lambda^*\in\partial h^*$, and remembering equation (1.10), we have, using (A.23)

$$\begin{aligned}
\psi^{(k_j)}(d^{(k_j)}) &= f(x^{(k_j)})+g^{\mathrm{T}}(x^{(k_j)})d^{(k_j)}+\tfrac{1}{2}d^{(k_j)\mathrm{T}}\nabla^2 f(x^{(k_j)})d^{(k_j)}\\
&\quad+\tfrac{1}{2}d^{(k_j)\mathrm{T}}\sum_{i=1}^{l}\lambda_i^{(k_j)}\nabla^2 c_i(x^{(k_j)})d^{(k_j)}+h(c^{(k_j)}+A^{(k_j)}d^{(k_j)})\\
&\geq f(x^{(k_j)}+d^{(k_j)})+h(c^*)\\
&\quad+\sum_{i=1}^{l}\lambda_i^*[c_i^{(k_j)}+\nabla^{\mathrm{T}}c_i^{(k_j)}d^{(k_j)}-c_i^*+\tfrac{1}{2}d^{(k_j)\mathrm{T}}\nabla^2 c_i^{(k_j)}d^{(k_j)}]+\mathrm{o}(\|d^{(k_j)}\|^2)\\
&= \phi(x^*)+f(x^{(k_j)}+d^{(k_j)})-f(x^*)\\
&\quad+\sum_{i=1}^{l}\lambda_i^*[c_i(x^{(k_j)}+d^{(k_j)})-c_i(x^*)]+\mathrm{o}(\|d^{(k_j)}\|^2)\\
&= \phi(x^*)+\tfrac{1}{2}(x^{(k_j)}+d^{(k_j)}-x^*)^{\mathrm{T}}W^*(x^{(k_j)}+d^{(k_j)}-x^*)+\mathrm{o}(\|d^{(k_j)}\|^2)\\
&\geq \phi(x^*)+\tfrac{1}{4}d'^{\mathrm{T}}W^*d'\|d^{(k_j)}\|^2+\mathrm{o}(\|d^{(k_j)}\|^2)
\end{aligned} \tag{A.29}$$

for all sufficiently large k. (A.29) contradicts (A.21) (using (A.19)) since $d'^{\mathrm{T}}W^*d' > 0$, which proves that (1) is true.

The proof of (2) is similar to that of (1). Assuming that (2) is not true, there exist k_j $(j = 1, 2, \ldots)$ such that

$$\|x^{(k_j)} - x^*\| = o(\|\hat{d}^{(k_j)}\|), \tag{A.30}$$

and consequently, from (1),

$$\|d^{(k_j)}\| = o(\|\hat{d}^{(k_j)}\|). \tag{A.31}$$

Recalling $\|d^{(k_j)} + \hat{d}^{(k_j)}\| \leq \rho^{(k_j)}$, we have $\|x^{(k_j)} - x^*\| = o(\rho^{(k_j)})$, so it follows that

$$\begin{aligned}\hat{\psi}^{(k_j)}(\hat{d}^{(k_j)}) &\leq \hat{\psi}^{(k_j)}(0) = \psi^{(k_j)}(d^{(k_j)}) + O(\|d^{(k_j)}\|^2) \\ &\leq \psi^{(k_j)}(x^* - x^{(k_j)}) + O(\|d^{(k_j)}\|^2) \leq \phi(x^*) + O(\|x^{(k_j)} - x^*\|^2).\end{aligned} \tag{A.32}$$

Choosing a subsequence if necessary, we assume that

$$\hat{d}^{(k_j)} / \|\hat{d}^{(k_j)}\| \to d^I, \tag{A.33}$$

and consequently

$$(x^{(k_j)} + d^{(k_j)} + \hat{d}^{(k_j)} - x^*) / \|\hat{d}^{(k_j)}\| \to d^I. \tag{A.34}$$

Since (A.30)-(A.32) imply

$$\phi(x^{(k_j)} + d^{(k_j)} + \hat{d}^{(k_j)}) = \hat{\psi}^{(k_j)}(\hat{d}^{(k_j)}) + O(\|\hat{d}^{(k_j)}\|^2) \leq \phi(x^*) + O(\|\hat{d}^{(k_j)}\|^2), \tag{A.35}$$

it follows that

$$d^I \in G^*. \tag{A.36}$$

Therefore

$$(x^{(k_j)} + d^{(k_j)} + \hat{d}^{(k_j)} - x^*)^T W^* (x^{(k_j)} + d^{(k_j)} + \hat{d}^{(k_j)} - x^*) \geq \tfrac{1}{2}(d^I)^T W^* d^I \|\hat{d}^{(k_j)}\|^2, \tag{A.37}$$

for all large j when $(d^I)^T W^* d^I > 0$. As in (A.29), we have

$$\begin{aligned}\hat{\psi}^{(k_j)}(\hat{d}^{(k_j)}) &\geq \phi(x^*) + o(\|\hat{d}^{(k_j)}\|^2) \\ &\quad + \tfrac{1}{2}(x^{(k_j)} + d^{(k_j)} + \hat{d}^{(k_j)} - x^*)^T W^* (x^{(k_j)} + d^{(k_j)} + \hat{d}^{(k_j)} - x^*) \\ &\geq \phi(x^*) + \tfrac{1}{4}(d^I)^T W^* d^I \|\hat{d}^{(k_j)}\|^2 + o(\|\hat{d}^{(k_j)}\|^2),\end{aligned} \tag{A.38}$$

for all large j, which contradicts (A.32) (using (A.30)). Therefore (2) is true.

We now prove (3). By (1), there exists $M_1 \geq 1$ such that

$$\|d^{(k)}\| \leq M_1 \|x^{(k)} - x^*\| \tag{A.39}$$

for all k. Noticing that $h(\cdot)$ is a polyhedral convex function, that $\{W^{(k)}\}$ is uniformly bounded, and that $\psi^{(k)}(\cdot)$ is defined by (1.3), we have that

$$\begin{aligned}\psi^{(k)}(\mathbf{0})-\min_{\|d\|\leqslant\alpha}\psi^{(k)}(d)&\geqslant\psi^{(k)}(\mathbf{0})-\min_{\|d\|\leqslant\alpha}[\psi^{(k)}(d)-\tfrac{1}{2}d^{\mathrm{T}}W^{(k)}d+\tfrac{1}{2}M_0\|d\|^2]\\&\geqslant\max_{0\leqslant t\leqslant 1}[t\Lambda_\alpha^k-\tfrac{1}{2}M_0\alpha^2t^2]\\&\geqslant\tfrac{1}{2}\min[\Lambda_\alpha^k,(\Lambda_\alpha^k)^2/M_0\alpha^2]\end{aligned}\tag{A.40}$$

for all $\alpha>0$, where M_0 is an upper bound on $\{|d^{\mathrm{T}}W^{(k)}d|:\|d\|=1\}$, and where

$$\begin{aligned}\Lambda_\alpha^k&=\psi^{(k)}(\mathbf{0})-\min_{\|d\|\leqslant\alpha}[\psi^{(k)}(d)-\tfrac{1}{2}d^{\mathrm{T}}W^{(k)}d]\\&=\psi^{(k)}(\mathbf{0})-\min_{\|d\|\leqslant\alpha}[f(x^{(k)})+g^{(k)\mathrm{T}}d+h(c^{(k)}+A^{(k)}d)].\end{aligned}\tag{A.41}$$

Since for a general convex function $F(\cdot)$,

$$\max_{\|d\|\leqslant\alpha}[F(x)-F(x+d)]\geqslant\min\left[1,\frac{\alpha}{\beta}\right]\times\max_{\|d\|\leqslant\beta}[F(x)-F(x+d)]$$

holds for any $\alpha,\beta\geqslant 0$, the convexity of $\psi^{(k)}(d)-\frac{1}{2}d^{\mathrm{T}}W^{(k)}d$ gives

$$\Lambda_{\|d^{(k)}\|}^k\geqslant\min[1,\|d^{(k)}\|/\|x^{(k)}-x^*\|]\Lambda_{\|x^{(k)}-x^*\|}^k,$$

consequently Lemmas 2.1, Lemma 2.2 (using $\alpha_0=\frac{1}{2}$), (A.39), and the fact that $M_0\geqslant 1$ give the following inequality, which holds for all large k,

$$\Lambda_{\|d^{(k)}\|}^k\geqslant\frac{c_1}{2M_1}\|x^{(k)}-x^*\|\,\|d^{(k)}\|.\tag{A.42}$$

It follows from (A.40) and (A.42) that there exists $M_2>0$ such that

$$\psi^{(k)}(\mathbf{0})-\psi^{(k)}(d^{(k)})\geqslant\frac{c_1}{4M_1}\min\left[1,\frac{c_1\|x^{(k)}-x^*\|}{2M_0M_1\|d^{(k)}\|}\right]\|x^{(k)}-x^*\|\,\|d^{(k)}\|\tag{A.43}$$

for all $k\geqslant M_2$. Thus (3) follows from (A.43) and (A.39).

We now prove (4). For all sufficiency large k, since

$$c(x^{(k)}+d^{(k)})-c(x^{(k)})-A^{(k)}d^{(k)}=\mathrm{O}(\|d^{(k)}\|^2),\tag{A.44}$$

and since $\mathrm{Rank}(A^{(k)})=m$ for all large k, there exists $\tilde{d}^{(k)}$ such that

$$c(x^{(k)}+d^{(k)})-c(x^{(k)})-A^{(k)}d^{(k)}=A^{(k)}\tilde{d}^{(k)}\tag{A.45}$$

and

$$\|\tilde{d}^{(k)}\|=\mathrm{O}(\|d^{(k)}\|^2).\tag{A.46}$$

Define

$$\beta_k=\max[\|d^{(k)}\|,\|d^{(k)}+\hat{d}^{(k)}\|],\tag{A.47}$$

then by the definition (1.9) of $r_e^{(k)}$, it follows that

$$r_e^{(k)} = \frac{1}{\Delta\psi^{(k)}}\Big\{\phi(x^{(k)}) - \phi(x^{(k)}+d^{(k)}) + q^{(k)}(d^{(k)}) + h(c(x^{(k)}+d^{(k)}))$$

$$- \min_{\|d^{(k)}+d\|\le\beta_k} [q^{(k)}(d^{(k)}+d) + h(c(x^{(k)}+d^{(k)}) + A^{(k)}d)]\Big\}$$

$$= \frac{1}{\Delta\psi^{(k)}}\Big\{\phi(x^{(k)}) + q^{(k)}(d^{(k)}) - f(x^{(k)}+d^{(k)})$$

$$- \min_{\|d^{(k)}+d\|\le\beta_k} [q^{(k)}(d^{(k)}+d) + h(c^{(k)} + A^{(k)}(d+d^{(k)}+\tilde{d}^{(k)}))]\Big\}$$

$$= \frac{1}{\Delta\psi^{(k)}}\Big[\phi(x^{(k)}) + q^{(k)}(d^{(k)}) - f(x^{(k)}+d^{(k)})$$

$$+ g^{(k)\mathrm{T}}\tilde{d}^{(k)} - \min_{\|d-\tilde{d}^{(k)}\|\le\beta_k} \psi^{(k)}(d) + \mathrm{o}(\|d^{(k)}\|^2)\Big]. \tag{A.48}$$

By the definition (1.5) of $q^{(k)}(\cdot)$ we have that

$$q^{(k)}(d^{(k)}) - f(x^{(k)}+d^{(k)}) = \tfrac{1}{2}d^{(k)\mathrm{T}} \sum_{i=1}^{m} \lambda_i^{(k)}\nabla^2 c_i^{(k)} d^{(k)} + \mathrm{o}(\|d^{(k)}\|^2), \tag{A.49}$$

and from (1.10), (1.13), (A.45) and (A.46), it follows that

$$g^{(k)\mathrm{T}}\tilde{d}^{(k)} = -\sum_{i=1}^{m} \lambda_i^{(k)}\nabla^{\mathrm{T}} c_i^{(k)}\tilde{d}^{(k)} + \mathrm{o}(\|d^{(k)}\|^2)$$

$$= -\sum_{i=1}^{m} \lambda_i^{(k)}[c_i(x^{(k)}+d^{(k)}) - c_i(x^{(k)}) - \nabla^{\mathrm{T}} c_i^{(k)} d^{(k)}] + \mathrm{o}(\|d^{(k)}\|^2)$$

$$= -\tfrac{1}{2}d^{(k)\mathrm{T}} \sum_{i=1}^{m} \lambda_i^{(k)}\nabla^2 c_i^{(k)} d^{(k)} + \mathrm{o}(\|d^{(k)}\|^2). \tag{A.50}$$

Hence from (A.47)-(A.50) and (3), noticing $\beta_k \le \rho^{(k)}$, we have that

$$r_e^{(k)} = \frac{\max_{\|d-\tilde{d}^{(k)}\|\le\beta_k} [\psi^{(k)}(0) - \psi^{(k)}(d) + \mathrm{o}(\|d^{(k)}\|^2)]}{\Delta\psi^{(k)}}$$

$$= \frac{\max_{\|d-\tilde{d}^{(k)}\|\le\beta_k} [\psi^{(k)}(0) - \psi_{(k)}(d)]}{\max_{\|d\|\le\beta_k} [\psi^{(k)}(0) - \psi^{(k)}(d)]} + \mathrm{o}(\|d^{(k)}\|)/\|d^{(k)}\|. \tag{A.51}$$

For each k we define

$$\Delta_\beta^k = \max_{\|d\|\le\beta} [\psi^{(k)}(0) - \psi^{(k)}(d)], \tag{A.52}$$

which is a monotonically increasing function of positive β and for any $\beta_0 > \beta > 0$ we let d_0 satisfy

$$\Delta_{\beta_0}^k = \psi^{(k)}(0) - \psi^{(k)}(d_0), \quad \|d_0\| \le \beta_0. \tag{A.53}$$

Thus, by the convexity of h and the definition of $\psi^{(k)}(\cdot)$, we have that

$$\begin{aligned}
\Delta_\beta^k &\geq \psi^{(k)}(\mathbf{0}) - \psi^{(k)}\left(\frac{\beta}{\beta_0}\,\boldsymbol{d}_0\right) \\
&\geq \frac{\beta}{\beta_0}\left[\psi^{(k)}(\mathbf{0}) - (\psi^{(k)}(\boldsymbol{d}_0) - \tfrac{1}{2}\boldsymbol{d}_0^{\mathrm{T}} W^{(k)} \boldsymbol{d}_0)\right] - \tfrac{1}{2}\left(\frac{\beta}{\beta_0}\right)^2 \boldsymbol{d}_0^{\mathrm{T}} W^{(k)} \boldsymbol{d}_0 \\
&= \frac{\beta}{\beta_0}\,\Delta_{\beta_0}^k + \tfrac{1}{2}\,\frac{\beta(\beta_0 - \beta)}{\beta_0^2}\,\boldsymbol{d}_0^{\mathrm{T}} W^{(k)} \boldsymbol{d}_0 \\
&= \frac{\beta}{\beta_0}\,\Delta_{\beta_0}^k\left[1 + \mathrm{O}\left(\frac{\beta_0 - \beta}{\beta_0}\right)\right],
\end{aligned} \tag{A.54}$$

where the last line depends on (3). Using (A.54) and (A.51), we deduce that

$$\begin{aligned}
r_e^{(k)} &\leq \frac{\Delta_{\beta_k}^k + \|\tilde{\boldsymbol{d}}^{(k)}\|}{\Delta_{\beta_k}^k} + \mathrm{o}(1) \\
&\leq \frac{\beta_k + \|\tilde{\boldsymbol{d}}^{(k)}\|}{\beta_k[1 + \mathrm{O}(\|\tilde{\boldsymbol{d}}^{(k)}\|/\beta_k)]} + \mathrm{o}(1) = 1 + \mathrm{o}(1),
\end{aligned} \tag{A.55}$$

and

$$\begin{aligned}
r_e^{(k)} &\geq \frac{\Delta_{\beta_k}^k - \|\tilde{\boldsymbol{d}}^{(k)}\|}{\Delta_{\beta_k}^k} + \mathrm{o}(1) \\
&\geq \frac{\beta_k - \|\tilde{\boldsymbol{d}}^{(k)}\|}{\beta_k}[1 + \mathrm{O}(\|\tilde{\boldsymbol{d}}^{(k)}\|/\beta_k)] + \mathrm{o}(1) = 1 + \mathrm{o}(1).
\end{aligned} \tag{A.56}$$

Therefore (4) follows.

Finally we prove (5). If (2.4) is not true, there exist $\alpha_1 \in (0, 1)$, k_j $(j = 1, 2, \ldots)$ such that $r^{(k_j)} \leq \alpha_1$ and

$$\|\boldsymbol{d}^{(k_j)}\| = \mathrm{O}(\|\hat{\boldsymbol{d}}^{(k_j)}\|). \tag{A.57}$$

To simplify notation, we denote k_j by j. Since $\{k_j\}$ is a subsequence of $\{k\}$, without loss of generality, we assume that

$$\boldsymbol{d}^{(j)}/\|\boldsymbol{d}^{(j)}\| \to \boldsymbol{d}', \qquad \hat{\boldsymbol{d}}^{(j)}/\|\hat{\boldsymbol{d}}^{(j)}\| \to \boldsymbol{d}^{I}. \tag{A.58}$$

Since our continuity assumption shows that

$$\psi^{(j)}(\mathbf{0}) - \psi^{(j)}(\boldsymbol{d}^{(j)}) = \phi(\boldsymbol{x}^{(j)}) - \phi(\boldsymbol{x}^{(j)} + \boldsymbol{d}^{(j)}) + \mathrm{O}(\|\boldsymbol{d}^{(j)}\|^2),$$

the definition (1.6) of $r^{(j)}$ and the assumption $r^{(j)} \leq \alpha_1$ ensure that

$$\psi^{(j)}(\mathbf{0}) - \psi^{(j)}(\boldsymbol{d}^{(j)}) = \mathrm{O}(\|\boldsymbol{d}^{(j)}\|^2). \tag{A.59}$$

Consequently the bounds (A.59) and (A.43) imply

$$\|\boldsymbol{x}^{(j)} - \boldsymbol{x}^*\| = \mathrm{O}(\|\boldsymbol{d}^{(j)}\|). \tag{A.60}$$

It follows from Lemma 2.2, (A.41), (A.60), the fact that (A.40) gives the condition

$$\Lambda^{k}_{\|d^{(k)}\|} \leqslant \max[2\Delta\psi^{(k)}, (2M_0\Delta\psi^{(k)}\|d^{(k)}\|^2)^{1/2}], \tag{A.61}$$

(A.59) and (1) that

$$\begin{aligned} \phi(x^{(j)}) - \phi(x^*) &\leqslant \frac{1}{\alpha_0} \Lambda^{j}_{\|x^{(j)}-x^*\|} \leqslant \frac{1}{\alpha_0} \Lambda^{j}_{\|d^{(j)}\|} \max\left[1, \frac{\|x^{(j)}-x^*\|}{\|d^{(j)}\|}\right] \\ &= O(\Lambda^{j}_{\|d^{(j)}\|}) = O(\|d^{(j)}\|^2) = O(\|x^{(j)}-x^*\|^2). \end{aligned} \tag{A.62}$$

Consequently

$$\begin{aligned} \phi(x^{(j)}+d^{(j)}) &= \psi^{(j)}(d^{(j)}) + O(\|d^{(j)}\|^2) \\ &\leqslant \phi(x^{(j)}) + O(\|d^{(j)}\|^2) \leqslant \phi(x^*) + O(\|d^{(j)}\|^2) \end{aligned} \tag{A.63}$$

Without loss of generality, we assume that

$$(x^{(j)} + d^{(j)} - x^*)/\|d^{(j)}\| \to \hat{d}'. \tag{A.64}$$

From (A.63) and (A.64), if $\hat{d}' \neq 0$, $\hat{d}' \in G^*$. Therefore for both zero and nonzero values of $\hat{d}'$,

$$d' \in G^* \tag{A.65}$$

since, due to (A.62), $\hat{d}' - d' \in G^*$. Similarly, it can be shown that

$$d^I \in G^*. \tag{A.66}$$

From (1.7), (1.5), (A.45), and (A.46), we deduce that

$$\begin{aligned} \hat{\psi}^{(j)}(d) &= q^{(j)}(d^{(j)} + d) + h(c^{(j)} + A^{(j)}(d + d^{(j)} + \tilde{d}^{(j)})) \\ &= \psi^{(j)}(d + d^{(j)} + \tilde{d}^{(j)}) - g^{(j)\mathrm{T}}\tilde{d}^{(j)} + o(\|d^{(j)}\|^2) + O(\|d\|\,\|d^{(j)}\|^2). \end{aligned} \tag{A.67}$$

Consequently, it follows from (A.54) that

$$\begin{aligned} &\hat{\psi}^{(j)}(-d^{(j)} - \tilde{d}^{(j)}) - \hat{\psi}^{(j)}(\hat{d}^{(j)}) \\ &= \max_{\|d+d^{(j)}\| \leqslant \beta_j} [\hat{\psi}^{(j)}(-d^{(j)} - \tilde{d}^{(j)}) - \hat{\psi}^{(j)}(d))] \\ &= \max_{\|d+d^{(j)}\| \leqslant \beta_j} [\psi^{(j)}(0) - \psi^{(j)}(d + d^{(j)} + \tilde{d}^{(j)})] + o(\|d^{(j)}\|^2) \\ &\geqslant \max_{\|d\| \leqslant \beta_j - \|\tilde{d}^{(j)}\|} [\psi^{(j)}(0) - \psi^{(j)}(d)] + o(\|d^{(j)}\|^2) \\ &\geqslant \psi^{(j)}(0) - \psi^{(j)}(d^{(j)}) + o(\|d^{(j)}\|^2) \\ &\geqslant \hat{\psi}^{(j)}(-d^{(j)} - \tilde{d}^{(j)}) - \hat{\psi}(-\tilde{d}^{(j)}) + o(\|d^{(j)}\|^2). \end{aligned} \tag{A.68}$$

Thus,

$$\hat{\psi}^{(j)}(\hat{d}^{(j)}) \leqslant \hat{\psi}^{(j)}(-\tilde{d}^{(j)}) + o(\|d^{(j)}\|^2). \tag{A.69}$$

Let $\Psi^{(j)}$ be defined by

$$\psi^{(j)}(\boldsymbol{d}+\boldsymbol{d}^{(j)}) = \Psi^{(j)}(\boldsymbol{d}) + \tfrac{1}{2}\boldsymbol{d}^{\mathrm{T}} W^{(k)} \boldsymbol{d}, \tag{A.70}$$

and note that $\Psi^{(j)}(\cdot)$ is a polyhedral convex function. Since $\psi^{(j)}(\boldsymbol{d})$ attains its minimum on $\{\boldsymbol{d}: \|\boldsymbol{d}\| \leq \rho^{(j)}\}$ at $\boldsymbol{d}^{(j)}$,

$$\Psi^{(j)}(\boldsymbol{d}) \geq \Psi^{(j)}(\boldsymbol{0}) \tag{A.71}$$

for all $\boldsymbol{d}$ satisfying

$$\|\boldsymbol{d}+\boldsymbol{d}^{(j)}\| \leq \rho^{(k)}. \tag{A.72}$$

In particular, if

$$\|\boldsymbol{d}^{(j)}+\hat{\boldsymbol{d}}^{(j)}+\tilde{\boldsymbol{d}}^{(j)}\| \leq \rho^{(k)}, \tag{A.73}$$

then

$$\Psi^{(j)}(\hat{\boldsymbol{d}}^{(j)}+\tilde{\boldsymbol{d}}^{(j)}) \geq \Psi^{(j)}(\boldsymbol{0}) = \psi^{(j)}(\boldsymbol{d}^{(j)}). \tag{A.74}$$

Consequently, it follows from (A.67), (A.70), (A.74) and (A.66) that

$$\begin{aligned}
\hat{\psi}^{(j)}(\hat{\boldsymbol{d}}^{(j)}) - \hat{\psi}^{(j)}(-\tilde{\boldsymbol{d}}^{(j)}) &= \psi^{(j)}(\boldsymbol{d}^{(j)}+\hat{\boldsymbol{d}}^{(j)}+\tilde{\boldsymbol{d}}^{(j)}) - \psi^{(j)}(\boldsymbol{d}^{(j)}) + \mathrm{o}(\|\boldsymbol{d}^{(j)}\|^2) \\
&\geq \tfrac{1}{2}(\hat{\boldsymbol{d}}^{(j)}+\tilde{\boldsymbol{d}}^{(j)})^{\mathrm{T}} W^{(j)} (\hat{\boldsymbol{d}}^{(j)}+\tilde{\boldsymbol{d}}^{(j)}) + \mathrm{o}(\|\boldsymbol{d}^{(j)}\|^2) \\
&\geq \tfrac{1}{4}(\boldsymbol{d}^{I})^{\mathrm{T}} W^{*} \boldsymbol{d}^{I} \|\boldsymbol{d}^{(j)}\|^2 + \mathrm{o}(\|\boldsymbol{d}^{(j)}\|^2)
\end{aligned} \tag{A.75}$$

for all large j, which, due to (A.57), contradicts (A.69). Therefore we assume that

$$\|\boldsymbol{d}^{(j)}+\hat{\boldsymbol{d}}^{(j)}+\tilde{\boldsymbol{d}}^{(j)}\| > \rho^{(j)} \tag{A.76}$$

for all j. Remembering (A.57) and that $\{j\}$ is a subsequence of $\{k\}$, we assume without loss of generality that

$$(\boldsymbol{d}^{(j)}+\hat{\boldsymbol{d}}^{(j)})/\|\hat{\boldsymbol{d}}^{(j)}\| \to \hat{\boldsymbol{d}}^{I}, \tag{A.77}$$

$$\hat{\boldsymbol{d}}^{I} \in G^{*}. \tag{A.78}$$

Define

$$\bar{\boldsymbol{d}}^{(j)} = (\boldsymbol{d}^{(j)}+\hat{\boldsymbol{d}}^{(j)}+\tilde{\boldsymbol{d}}^{(j)})(1-2\|\tilde{\boldsymbol{d}}^{(j)}\|/\rho^{(j)}). \tag{A.79}$$

Then, it follows that for sufficiently large j

$$\|\bar{\boldsymbol{d}}^{(j)}\| < \rho^{(j)}, \qquad (\bar{\boldsymbol{d}}^{(j)}-\boldsymbol{d}^{(j)})/\|\hat{\boldsymbol{d}}^{(j)}\| \to \boldsymbol{d}^{I}, \tag{A.80}$$

since $\|\boldsymbol{d}^{(j)}+\hat{\boldsymbol{d}}^{(j)}+\tilde{\boldsymbol{d}}^{(j)}\| \leq \rho^{(j)} + \|\tilde{\boldsymbol{d}}^{(j)}\|$. Similarly to (A.75), it can be shown that

$$\hat{\psi}^{(j)}(\bar{\boldsymbol{d}}^{(j)}-\boldsymbol{d}^{(j)}-\tilde{\boldsymbol{d}}^{(j)}) - \hat{\psi}^{(j)}(-\tilde{\boldsymbol{d}}^{(j)}) \geq \tfrac{1}{4}(\boldsymbol{d}^{I})^{\mathrm{T}} W^{*} \boldsymbol{d}^{I} \|\boldsymbol{d}^{(j)}\|^2 + \mathrm{o}(\|\boldsymbol{d}^{(j)}\|^2). \tag{A.81}$$

Now (A.79), (A.77), (A.46) and (A.57) imply that

$$\bar{\boldsymbol{d}}^{(j)}/\|\hat{\boldsymbol{d}}^{(j)}\| \to \hat{\boldsymbol{d}}^{I}. \tag{A.82}$$

Using (A.67), (A.46), and the continuity of $\psi^{(j)}(d)-\frac{1}{2}d^{\mathrm{T}}W^{(j)}d$, we deduce that

$$\begin{aligned}
&\hat{\psi}^{(j)}(-d^{(j)}-\tilde{d}^{(j)})-\hat{\psi}^{(j)}(\bar{d}^{(j)}-d^{(j)}-\tilde{d}^{(j)})\\
&=\psi^{(j)}(\mathbf{0})-\psi^{(j)}(\bar{d}^{(j)})+o(\|d^{(j)}\|^2)\\
&\geqslant(1-2\|\tilde{d}^{(j)}\|/\rho^{(j)})[\psi^{(j)}(\mathbf{0})-\psi^{(j)}(d^{(j)}+\hat{d}^{(j)}+\tilde{d}^{(j)})]+o(\|d^{(j)}\|^2)\\
&=\psi^{(j)}(\mathbf{0})-\psi^{(j)}(d^{(j)}+\hat{d}^{(j)}+\tilde{d}^{(j)})+o(\|d^{(j)}\|^2)\\
&=\hat{\psi}^{(j)}(-d^{(j)}-\tilde{d}^{(j)})-\hat{\psi}^{(j)}(\hat{d}^{(j)})+o(\|d^{(j)}\|^2)
\end{aligned}\tag{A.83}$$

(A.81) and (A.83) show that (A.75) holds, which contradicts (A.69). The contradiction shows that (2.4) is true.

Remembering our notation that $\hat{c}^{(k)}=c(x^{(k)}+d^{(k)})$, by the definition (1.7) of $\hat{\psi}^{(k)}(\cdot)$, we have

$$\begin{aligned}
\Delta\hat{\psi}^{(k)}&=q^{(k)}(d^{(k)})+h(\hat{c}^{(k)})-[q^{(k)}(d^{(k)}+\hat{d}^{(k)})+h(\hat{c}^{(k)}+A^{(k)}\hat{d}^{(k)})]\\
&=f(\hat{x}^{(k)})+h(\hat{c}^{(k)})+\tfrac{1}{2}d^{(k)\mathrm{T}}\sum_{i=1}^{m}\lambda_i^{(k)}\nabla^2c_i^{(k)}d^{(k)}-[f(\tilde{x}^{(k)})+h(c(\tilde{x}^{(k)}))\\
&\quad+\tfrac{1}{2}(d^{(k)}+\hat{d}^{(k)})^{\mathrm{T}}\sum_{i=1}^{m}\lambda_i^{(k)}\nabla^2c_i^{(k)}(d^{(k)}+\hat{d}^{(k)})\\
&\quad+o(\|d^{(k)}\|^2)+O(\|d^{(k)}+\hat{d}^{(k)}\|^2)+O(\|d^{(k)}\|\,\|\hat{d}^{(k)}\|)+O(\|\hat{d}^{(k)}\|^2).\\
&=\phi(x^{(k)}+d^{(k)})-\phi(x^{(k)}+d^{(k)}+\hat{d}^{(k)})\\
&\quad+o(\|d^{(k)}\|^2)+O(\|d^{(k)}\|\,\|\hat{d}^{(k)}\|)+O(\|\hat{d}^{(k)}\|^2).
\end{aligned}\tag{A.84}$$

Hence (2.5) follows from (1.9), (2.4), (3) and (4). This completes the proof of Lemma 2.3. □

References

[1] R.H. Bartels, A.R. Conn and J.W. Sinclair, "Minimization techniques for piecewise differentiable functions: the L_1 solution to an overdetermined linear system", *SIAM Journal of Numerical Analysis* 15 (1978) 224-241.

[2] R. Fletcher, "A Fortran subroutine for general quadratic programming", A.E.R.E. Report 6370, Harwell, England (1970).

[3] R. Fletcher, "A model algorithm for composite NDO problem", *Mathematical Programming Studies* 17 (1982) 67-76.

[4] R. Fletcher, *Practical methods of optimization Vol. 2, Constrained optimization* (John Wiley, New York, 1981).

[5] R. Fletcher, "Second order correction for nondifferentiable optimization", in: G.A. Watson, ed., *Numerical analysis* (Springer-Verlag, Berlin, 1982) pp. 85-114.

[6] D. Goldfarb and A. Idnani, "A numerically stable dual method for solving strictly convex quadratic programs", *Mathematical Programming* 27 (1983) 1-33.

[7] P.E. Gill and W. Murray, *Numerical methods for constrained optimization* (Academic Press, London, 1974).

[8] P.E. Gill and W. Murray, "Numerically stable methods for quadratic programming", *Mathematical Programming* 14 (1978) 349-372.

[9] P.E. Gill, W. Murray, M.A. Saunders and M.H. Wright, "Linearly constrained optimization", in: M.J.D. Powell, ed., *Nonlinear optimization* 1981 (Academic Press, London, 1982).
[10] W. Murray and M.L. Overton, "A projected Lagrangian algorithm for nonlinear minmax optimization", *SIAM Journal on Scientific and Statistical Computing* 1 (1980) 345-370.
[11] W. Murray and M.L. Overton, "A projected Lagrangian algorithm for nonlinear L_1 optimization", *SIAM Journal on Scientific and Statistical Computing* 2 (1981) 207-224.
[12] J. Nocedal and M. Overton, "Projected Hessian updating algorithms for nonlinear constrained optimization", Report 59, Computer Science Department, New York University (November, 1983).
[13] M.J.D. Powell, "A new algorithm for unconstrained optimization", in: J.B. Rosen, O.L. Mangasarian and K. Ritter, eds., *Nonlinear programming* (Academic Press, New York, 1970).
[14] M.J.D. Powell, "Convergence properties of a class of minimization algorithms", in: O.L. Mangasarian, R.R. Meyer and S.M. Robinson, eds., *Nonlinear programming* 2 (Academic Press, New York, 1975).
[15] Y. Yuan, "Some properties of trust region algorithms for nonsmooth optimization", Report DAMTP 1983/NA4, University of Cambridge.
[16] Y. Yuan, "An example of only linear convergence of trust region algorithms for nonsmooth optimization", *IMA Journal of Numerical Analysis* 4 (1984) 327-335.