A flexible framework for cubic regularization algorithms for non-convex optimization in function space

Anton Schiela

April 4, 2014

Abstract

We propose a cubic regularization algorithm that is constructed to deal with non-convex minimization problems in function space. It allows for a flexible choice of the regularization term and thus accounts for the fact that in such problems one often has to deal with more than one norm. Global and local convergence results are established in a general framework.

AMS MSC 2000: 74B20, 49M37

Keywords: non-convex optimization, optimization in function space, cubic regularization

1 Introduction

In broad terms, non-linear optimization algorithms rely on two types of models. A local model of the functional f that can be treated with techniques of linear algebra, and a rough model for the error of the local model. State of the art optimization methods usually employ a quadratic model for f and a rough parametrized model for remaining difference between the quadratic model and the functional, i.e., the local error. Usually such an error model is based on a norm $\|\cdot\|$. In classical trust region methods (cf. e.g. [4]) the error model is 0 inside a ball of varying radius around the current iterate and $+\infty$ otherwise. In cubic regularization methods the error model is chosen according to the assumption that the difference of f and its quadratic model is of third order. Thus, a scaling of $\|\cdot\|^3$ is taken as a model for the error.

The reason for introducing such an error model (in contrast to a line-search approach) is the wish to transfer information about the actual error attained at sampling points (i.e., at trial corrections) to a whole neighborhood of the current iterate. The implicit assumption behind this reasoning is that the error indeed behaves more or less isotropically with respect to the chosen norm. One assumes that the obtained sampled information is indeed representative for the error in the neighborhood. Of course, this cannot be guaranteed in general, but in those cases where the error model predicts the actual error well, we expect a stable and efficient behavior of our algorithm. Thus, in particular in large scale problems, it pays off to choose the error model, and thus the underlying norm, carefully.

In this paper we consider a cubic error model. This idea is not new, and, to the best knowledge of the authors, has first been proposed by Griewank [9] in an unpublished technical report. Independently, Weiser, Deuflhard, and Erdmann [16] proposed a cubic regularization in an attempt to generalize the works [5, 6] on convex optimization to the non-convex case. Focus was laid on the construction of estimates for the third order remainder term. Even more recently Cartis, Gould, and Toint proposed an algorithmic framework, similar to trust-region methods, but with a cubic regularization term, and provided detailed first and second order convergence analysis [2] and a complexity analysis [3]. Common idea of all these methods is the cubic regularization, but apart from this basic idea, the proposed methods differ significantly.

An important class of large scale optimization problems comes from discretizations of problems with partial differential equations. These may comprise problems of energy minimization, such as nonlinear elliptic problems, or problems from optimal control of partial differential equations. If one wants to apply cubic regularization methods in this setting, one is naturally led to versions of these methods that work in function space, and the need for a convergence theory in function space arises. This can be done in a relatively straightforward way, if one chooses the framework proposed, e.g., by [16] or [2].

Such a straightforward generalization, however, would ignore one of the main strengths that lie in the functional analytic treatment of such problems, namely the possibility to work with more than one norm in order to capture the main features of the problem at hand. Different choices can be made concerning boundedness of first and second derivatives, and the limiting behavior of remainder terms. For example, second derivatives of non-convex functionals are often bounded from below with respect to a strictly weaker norm, than they are bounded from above. This is, because well posed optimization problems in function space usually exhibit a subtle analytic structure, consisting of a combination of convexity and compactness. Usually a careful analysis of the problem at hand reveals the right norms to be chosen.

The concept of cubic regularization allows to exploit the insights, gained from such an analysis if one is willing to exchange the usual $\|\cdot\|^3$ term by a more general third order functional. This will lead to SQP methods that use an error model that can be better adapted to the problem at hand, and thus provide additional flexibility for problem adapted optimization methods.

The aim of this paper is to explore this idea and find an algorithmic and theoretic framework in function space for the flexible choice of cubic regularization terms. Our framework employs a weak norm $|\cdot|$ and a strong norm $\|\cdot\|$ which are used to formulate the required smoothness assumptions on f. Then conditions, depending on these norms, are imposed on the cubic regularization term which allow for a global and a local convergence analysis.

In particular, we will introduce our flexible analytic framework in Section 2, which is designed as a compromise between simplicity and generality. We discuss possible extensions and give some examples to illustrate the abstract concepts. In Section 3 we develop our algorithmic framework. It resembles in a couple of points the classical trust region-like algorithms with the usual fraction of model decrease acceptance criterion and a fraction of Cauchy decrease condition. However, the latter condition has to be modified to take into account non-equivalence of norms. Also here it was our aim to leave as much flexibility for concrete implementations of algorithms, concerning updates of regularization parameters and computation of steps. Within this framework we show in Section 4 global and local convergence results. The main challenge here was to identify the theoretically relevant quantities that have to be considered in order to show convergence.

The present paper represents a step into a new area of research, and we emphasize again that its focus is to establish a framework for algorithms, rather than propose concrete implementations. Concrete algorithms are subject to current research, but still go beyond the scope of the present paper. We postpone this and also the publication of numerical studies to a future publication.

2 Functional analytic framework

Consider for a given function $f : X \to \mathbb{R}$ on a linear space X the minimization problem

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

Suppose that we can compute for each $x \in X$ a quadratic model, consisting of a a linear functional $f'_x : X \to \mathbb{R}$ and a bilinear form $H_x : X \times X \to \mathbb{R}$:

$$q_x(\delta x) := f(x) + f'_x \delta x + \frac{1}{2} H_x(\delta x, \delta x). \tag{1}$$

Further, let us denote an upper bound on the error as follows

$$f(x + \delta x) - q_x(\delta x) \leq w_x(\delta x). \tag{2}$$

Later, we will impose various smoothness assumptions on f, i.e., make assumptions on the limiting behavior of w_x for small δx. Depending on the smoothness of the problem, $w_x(\lambda \delta x)$ may be of higher order locally, such as $o(\lambda)$, $o(\lambda^2)$ or even $O(\lambda^3)$ as $\lambda \to 0$. The latter case occurs, for example, if H is the second Fréchet derivative of f, and Lipschitz. This "generic case" motivates the construction of the following cubic model for f:

$$f(x + \delta x) - f(x) \approx m_x^\omega(\delta x) := f'_x \delta x + \frac{1}{2} H_x(\delta x, \delta x) + \frac{\omega}{6} R_x(\delta x). \tag{3}$$

Here R_x is a functional, which is homogenous of order 3:

$$R_x(\lambda \delta x) = |\lambda|^3 R_x(\delta x) \qquad \forall \lambda \in \mathbb{R} \tag{4}$$

and positive:

$$R_x(\delta x) > 0 \quad \forall \delta x \neq 0.$$

In (3) the parameter $\omega > 0$ is updated adaptively during the course of the algorithm in order to globalize the method. Comparison of (2) and (3) yields that $(\omega/6)R_x$ can be seen as a model for w_x.

If X is equipped with the norm $\|\cdot\|$, the classic cubic regularization method uses $R_x(\delta x) := \|\delta x\|^3$.

However, in most function space problems an adequate analysis requires the use of several non-equivalent norms. There are a couple of different issues, which each on its own may require a separate choice. This is why we aim for a theoretical framework that is flexible with respect to choosing more than one norm.

Assumptions for global convergence. Let us collect the following set of assumptions for later reference, which are needed to show global convergence, i.e., $\liminf_{k\to\infty} \|f'_{x_k}\| = 0$ for our algorithm. Among them the only *standing assumptions* are reflexivity of X and existence of f'_x in X^*. All other assumptions will be referenced later, when needed.

(i) Let $(X, \|\cdot\|)$ be a reflexive Banach space. The *primary norm* $\|\cdot\|$ on X has to be strong enough that f is continuously differentiable on X. This means that we have the inequality:

$$\|f'_x\| := \sup_{\|\delta x\|=1} |f'_x \delta x| < \infty,$$

and, moreover that $x_k \to x_*$ in $(X, \|\cdot\|)$ implies $f'_{x_k} \to f'_{x_*}$ in $(X, \|\cdot\|)^*$.

(ii) The *secondary norm* $|\cdot|$ of X is used to describe possible non-convexity of the quadratic model q_x. We do *not* assume completeness of $(X, |\cdot|)$, which allows to choose $|\cdot|$ strictly weaker than $\|\cdot\|$. With the help of our two norms we impose a condition of Gårding-type:

$$\exists \gamma > -\infty,\ \Gamma < \infty : \quad \gamma |v|^2 \le H_x(v,v) \le \Gamma \|v\|^2. \tag{5}$$

Hence, H_x is assumed to be bounded below in a different norm than it is bounded above. Similar conditions appear, for example, in the theory of pseudo-monotone operators. In the next section we will discuss some examples, where this condition is fulfilled.

(iii) The main purpose of R_x is to compensate the possible non-convexity of the quadratic subproblems and to model the remainder term. Thus, we impose the following flexible boundedness and coercivity condition (without a constant in the left inequality for simplicity):

$$|v|^3 \le R_x(v) \le C\|v\|^3 \qquad \forall v \in X \tag{6}$$

In view of (5) the left inequality guarantees that the cubic subproblems (3) are bounded from below. The right inequality is mainly needed only in directions of steepest descent and prevents Cauchy steps from becoming too short.

The chosen set of assumptions presents a compromise between generality and simplicity. It is slightly stronger than needed by our global convergence theory. For example, concerning smoothness of f we will only make use of the results of the following lemma. Later we will also discuss possible weakenings of (5) and (6) (cf. (35) and (36), below).

Lemma 2.1. *Consider a sequence $x_k \to x_*$ and a sequence δv_k. Then from Fréchet differentiability of f and continuity of f' at x_* and* (5) *it follows for the remainder term, defined in* (2)*:*

$$\lim_{k\to\infty} \|\delta v_k\| + \omega_k R_{x_k}(\delta v_k) = 0 \quad \Rightarrow \quad \lim_{k\to\infty} \frac{w_{x_k}(\delta v_k)}{\|\delta v_k\|} = 0. \tag{7}$$

Moreover, (7) *implies the following relation:*

$$\frac{\omega_k R_{x_k}(\delta v_k)}{\|\delta v_k\|} \ge W_0 > 0 \ \wedge\ \frac{w_{x_k}(\delta v_k)}{\|\delta v_k\|} \ge C_0 > 0 \quad \Rightarrow \quad \omega_k R_{x_k}(\delta v_k) \ge R_0 > 0. \tag{8}$$

Proof. Concerning (7), it follows from a standard result of analysis (cf. e.g. [12, Thm. 25.23], an application of the fundamental theorem of calculus) in combination with (5) that

$$\|\delta v_k\| \to 0 \quad \Rightarrow \quad \lim_{k\to\infty} \frac{w_{x_k}(\delta v_k)}{\|\delta v_k\|} = 0.$$

This property is also called *strict* or *strong* differentiability of f at x_*. Equation (7) is just a further weakening of that.

To show (8) we abbreviate $r_k := \omega_k R_{x_k}(\delta v_k)$ and $d_k := \|\delta v_k\|$. By assumption we conclude $r_k \ge W_0 d_k$ and via (7) that $r_k + d_k \ge D_0 > 0$. It follows

$$r_k \ge D_0 - d_k \ge D_0 - W_0^{-1} r_k \quad \Rightarrow \quad r_k \ge R_0 := \frac{D_0 W_0}{W_0 + 1} > 0.$$

□

Global convergence results usually rely on proof by contradiction. The following lemma, which uses reflexivity of X, will serve as a key facility to obtain this contradiction (see Lemma 4.3).

Lemma 2.2. *Let $x_k \in X$ be a sequence such that $|x_k| \to 0$ and $\|x_k\|$ is bounded. Then by reflexivity of X:*

$$x_k \rightharpoonup 0 \text{ weakly in } (X, \|\cdot\|).$$

Proof. Since $(X, \|\cdot\|)$ is reflexive, x_k has a weakly convergent subsequence, say $x_{k_j} \rightharpoonup x_*$. Since $|x_{k_j}| \to 0$ we conclude that for each $\varepsilon > 0$, x_{k_j} is eventually contained in a ball of $|\cdot|$-radius ε, and thus also x_*. It follows that $x_* = 0$. This also shows that every possible weak accumulation point of our sequence is 0, so by a standard argument the whole sequence converges weakly to 0. □

Remark 2.3. A slight relaxation of our reflexivity assumption on X is conceivable. Let ${}^*X$ be a separable Banach space and define X as its dual: $X = ({}^*X)^*$. Under these conditions Lemma 2.2 would still hold with weak convergence replaced by weak* convergence. However, to be able to apply this lemma later, we would have to impose the condition that the derivatives f'_{x_k} are all contained in the closed subspace ${}^*X$ of $X^* = ({}^*X)^{**}$. Well known examples for spaces X with these properties are $L_\infty(K)$, where ${}^*X = L_1(K)$ and spaces of measures $M(K)$, where ${}^*X = C(K)$.

Assumptions for fast local convergence. If we want to show fast local convergence we need the following additional assumptions, which strengthen assumption (i) and (ii) from the above list:

(i)$_{\mathrm{loc}}$ Setting $\delta x_k = x_{k+1} - x_k$ we need a second order approximation error estimate in (2):

$$\lim_{\|x_k - x_*\| \to 0} \frac{w_{x_k}(\delta x_k)}{\|\delta x_k\|^2} = 0, \tag{9}$$

close to a local minimizer, which is fulfilled in particular, if f is twice continuously differentiable and $H_x = f''_x$.

(ii)$_{\mathrm{loc}}$ Locally, we have to impose stronger assumptions on H_x. Close to a minimizer we assume in addition to (5) ellipticity of H_x with respect to the strong norm $\|\cdot\|$:

$$\exists \gamma > 0 : \quad \gamma \|\delta x\|^2 \le H_x(\delta x, \delta x). \tag{10}$$

Discussion. In contrast to convergence theory in finite dimensions, for the application of our framework to concrete problems the choice of function spaces and corresponding norms is of high interest. Let us thus discuss possible strategies for finding appropriate norms.

- If we are interested in global convergence only, we may choose the strong norm $\|\cdot\|$ as strong as we wish. This means, if our assumptions hold for some Banach space $(X, \|\cdot\|)$, then they also hold for a stronger norm, defined on a smaller Banach space, densely embedded into $(X, \|\cdot\|)$. However, the stronger $\|\cdot\|$ is chosen, the weaker is its dual norm, the weaker is our convergence result $\|f'_{x_k}\| \to 0$. Thus, we should choose $\|\cdot\|$ as weak as the analysis of our problem at hand allows. This means that $\|\cdot\|$ has to be strong enough to guarantee boundedness of derivatives, (35) and (7). In highly nonlinear problems (7) will probably be the most restrictive condition.

- The choice of the norm $\|\cdot\|$ has direct algorithmic consequences, because directions of quasi steepest descent (cf. Section 3.2) are defined via this norm. From a practical point of view, one should choose $\|\cdot\|$ such that those directions are not too expensive to compute. For example, for H^1-elliptic problems one can use a multigrid-preconditioner that induces a norm that is equivalent to $\|\cdot\|_{H^1}$. In any case, it is advantageous to choose $(X, \|\cdot\|)$ as a Hilbert space, if possible. In this case the direction of steepest descent can be computed by solving a *linear* operator equation. Otherwise, this computation is probably a non-linear problem.

- If fast local convergence is the aim, then a good choice for $\|\cdot\|$ can usually be found by analysis of f''_x under the assumption of second order sufficient optimality conditions (SSC). Our assumption (10) together with (5) corresponds to classical ellipticity and continuity assumptions with respect to $\|\cdot\|$. In other terms, H_x determines the choice of norm up to equivalence, and $v \to \sqrt{H_x(v,v)}$ itself defines an equivalent Hilbert space norm. Algorithmically, the latter is the optimal choice, since then $\Gamma = \gamma = 1$. An inadequately strong norm would choke off fast local convergence, because (10) would be unlikely to hold for some $\gamma > 0$.

- Very difficult problems, however, only allow for a weak form of the (SSC), including a so called *two-norm discrepancy* (see e.g. [13]). In that case it is not possible that a norm satisfying both (10) and (9) can be found. Rather, a stricter norm $|||\cdot|||$ has to be introduced and only the weaker condition

$$\lim_{|||x_k - x_*||| \to 0} \frac{w_{x_k}(\delta x_k)}{\|\delta x_k\|^2} = 0$$

  can be shown to hold. This class of problems is not amenable to fast local convergence. The problem is that although $\|x_k - x_*\| \to 0$ can be shown for a minimization method, the stronger property $|||x_k - x_*||| \to 0$ is not valid in general. Even more, one cannot even guarantee that $|||x_k - x_*|||$ is well defined for the iterates x_k.

- If f is convex, then γ can be set to 0 and thus $|\cdot|$ can be chosen arbitrarily weak, as long as it defines a norm on X. Nevertheless, R_x should be chosen to model the remainder terms appropriately.

2.1 Examples

To get a feeling for the peculiarities of this class of problems, we will discuss a couple of typical examples. This will help to understand the ideas behind our functional analytic framework. We restrict ourselves to simple settings and dispense with a deeper discussion in order to keep this section concise.

2.1.1 Two notorious toy examples

The following well known toy examples serve as an illustration, why the choice of two norms in (5) is quite natural in infinite dimensional optimization.

In contrast to finite dimensional problems, where existence of minimizers is usually easy to obtain, infinite dimensional problems are notoriously hard to analyse. The main reason is the *lack of compactness* of closed and bounded sets in infinite dimensions. By turning to weakly converging sequences, compactness can often be retained, but only in connection with *convex* problems. Still, certain classes of non-convex problems are tractable, as long as there are other, additional compactness results available, such as compact Sobolev embeddings, e.g., $E : H_0^1 \hookrightarrow L_4$.

Let us consider the following two functionals:

$$\phi(v) := \int_0^1 \frac{1}{2} v^2\, dx, \qquad \psi(v) := \int_0^1 (v^2-1)^2\, dx.$$

We observe that $\phi : L_2(\Omega) \to \mathbb{R}$ is convex, while $\psi : L_4(\Omega) \to \mathbb{R}$ is non-convex (with a w-shape) and both functionals are non-negative.

The following minimization problem, which involves the first derivative $u' = du/dx$ is well defined in $H_0^1(0,1)$:

$$\min_{u \in H_0^1(0,1)} f(u) := \phi(u') + \psi(Eu)$$

The non-convex part of this problem appears together with the compact Sobolev embedding E, which is the reason why this problem admits a minimizer.

In contrast, the following minimization problem, which is well defined on $W_0^{1,4}(0,1)$:

$$\min_{u \in W_0^{1,4}(0,1)} \tilde{f}(u) := \phi(Eu) + \psi(u')$$

does not admit a global minimizer. This can be seen by considering a sequence of functions u_k' , which only take the values -1 and $+1$ and oscillate between them with higher and higher frequency as $k \to \infty$. All these functions minimize ψ with $\psi(u_k') = 0$. Then u_k is a sequence of saw-tooth functions, which tends to $u_* = 0$ (in $L_2(0,1)$). This in turn implies $\phi(u_k) \to 0$, so that $\inf_k \phi(Eu_k) + \psi(u_k') = 0$. However, this infimum is not attained, because $\psi(u_*') = 1 > 0$.

In view of (5), let us consider the (formal) second derivatives of f and $\tilde{f}$, for example at $u_* = 0$:

$$f''_{u_*}(\delta u, \delta u) = \int_0^1 (\delta u')^2 - 4\delta u^2\, dx \quad \Rightarrow \quad -4\|\delta u\|^2_{L_2} \le f''_{u_*}(\delta u, \delta u) \le \|\delta u\|^2_{H^1}$$

$$\tilde{f}''_{u_*}(\delta u, \delta u) = \int_0^1 \delta u^2 - 4(\delta u')^2\, dx \quad \Rightarrow \quad -4\|\delta u\|^2_{H^1} \le \tilde{f}''_{u_*}(\delta u, \delta u) \le \|\delta u\|^2_{H^1}.$$

We observe that (5) is fulfilled with two different norms for f with the weaker norm measuring the non-convexity. We may set $|\cdot| = \|\cdot\|_{L_2}$ and $\|\cdot\| = \|\cdot\|_{H^1}$. For $\tilde{f}$ the choice of a weaker norm for the lower bound is not possible.

The bottom line is that compactness, the principle on which existence of minimizers for non-convex problems rests, is closely related to the presence of two norms in (5), where the lower bound is measured in a weaker norm than the upper bound.

2.1.2 Semi-linear elliptic PDEs

In the following let $\Omega \subset \mathbb{R}^d$ $(1 \le d \le 3)$ be a smoothly bounded open domain, and $x \in \Omega$ the spatial variable. Further, let $H_0^1(\Omega)$ be the usual Sobolev space of weakly differentiable functions on Ω with zero boundary conditions. By the Sobolev embedding theorem there is a continuous embedding $H_0^1(\Omega) \hookrightarrow L_6(\Omega)$ for $d \le 3$. Further, denote by $v \cdot w$ the euclidean scalar product of $v, w \in \mathbb{R}^d$.

As a prototypical example we consider the following energy functional of a nonlinear elliptic PDE:

$$f(u) := \int_\Omega \frac{1}{2} \nabla u(x) \cdot \nabla u(x) + a(u(x), x)\, dx.$$

Here $a(u,x)$ is a Carathéodory function that is twice continuously differentiable with respect to u.

We are looking for the solution of the minimization problem

$$\min_{u\in H_0^1(\Omega)} f(u).$$

Its (formal) first and second derivatives are given by:

$$\begin{aligned} f_u'\delta u &= \int_\Omega \nabla u\cdot\nabla\delta u + \frac{\partial}{\partial u}a(u,x)\delta u\,dx \\ f_u''(\delta u_1,\delta u_2) &= \int_\Omega \nabla\delta u_1\cdot\nabla\delta u_2 + \frac{\partial^2}{\partial u^2}a(u,x)\delta u_1\delta u_2\,dx \end{aligned}$$

Let us analyse these functionals. We may assume that $u\in H_0^1(\Omega)$, which implies that

$$\left|\int_\Omega \nabla u\cdot\nabla\delta u\,dx\right| \le c(u)\|\delta u\|_{H_0^1}$$

and similarly

$$\left|\int_\Omega \nabla\delta u_1\cdot\nabla\delta u_2\,dx\right| \le \|\delta u_1\|_{H_0^1}\|\delta u_2\|_{H_0^1}.$$

Under the assumption that $\frac{\partial}{\partial u}a(u,\cdot)\in L_{6/5}$ and $\frac{\partial^2}{\partial u^2}a(u,\cdot)\in L_{3/2}$, we obtain the following estimates for the second parts of the derivatives via the Hölder inequality:

$$\begin{aligned} \left|\int_\Omega \frac{\partial}{\partial u}a(u,x)\delta u\,dx\right| &\le \|\frac{\partial}{\partial u}a(u,\cdot)\|_{L_{6/5}}\|\delta u\|_{L_6} \le c(u)\|\delta u\|_{H_0^1} \\ \left|\int_\Omega \frac{\partial^2}{\partial u^2}a(u,x)\delta u_1\delta u_2\,dx\right| &\le \|\frac{\partial^2}{\partial u^2}a(u,\cdot)\|_{L_{3/2}}\|\delta u_1\|_{L_6}\|\delta u_2\|_{L_6} \end{aligned}$$

Taking these estimates together, we obtain the following results:

$$\begin{aligned} |f_u'\delta u| &\le c(u)\|\delta u\|_{H_0^1} \\ c_0(u)\|\delta u\|_{L_6}^2 \le f_u''(\delta u,\delta u) &\le c_1(u)\|\delta u\|_{H_0^1}^2, \end{aligned}$$

where $c_0(u)>-\infty$ may be negative, and $c(u),c_1(u)<+\infty$ are positive. Our first observation is that f' and f'' can be bounded (from above) via a strong norm $\|\cdot\| := \|\cdot\|_{H_0^1}$, while it only takes a weaker norm $|\cdot| := \|\cdot\|_{L_6}$ to formulate a lower bound on f''. The choice of the weaker norm can be taken differently. If, for example $\frac{\partial^2}{\partial u^2}a(u,\cdot)\in L_\infty$, then $|\cdot| := \|\cdot\|_{L_2}$ is sufficient.

2.1.3 Nonlinear optimal control: black-box approach

The aim in (PDE constrained) optimal control is to minimize a cost functional subject to a (partial) differential equation as equality constraint. For an introduction into this topic, we refer to the textbooks [11, 14, 10, 15]. Usually, the optimization variable is divided into a *control* u which enters the differential equation as data, and the *state* y, which is the corresponding solution. This relation can be described by a nonlinear operator via $y=S(u)$. Elimination of y then yields an optimization problem of the following form:

$$\min_{u\in U} f(S(u),u).$$

This general problem, however is hardly tractable theoretically, and thus, one restricts considerations often to the following special case:

$$f(S(u), u) = g_1(S(u)) + g_2(u) = g_1(S(u)) + \frac{\alpha}{2}\|u\|_U^2.$$

Here, the convexity of g_2 is crucial, but our special quadratic choice for it is a matter of simplicity.

If S is the solution operator for a non-linear elliptic PDE, and $U = L_2(\Omega)$, then an appropriate choice of norms would be

$$\|v\| := \|v\|_{L_2(\Omega)} \qquad |v| := \|S'(u)v\|_{H_0^1}.$$

Here $|\cdot|$ depends on u, an issue that is encountered frequently. We will ignore this, however, for the sake of simplicity. Usually, $S'(u)$ is a compact linear operator, so that $|\cdot|$ is strictly weaker than $\|\cdot\|$.

Due to the special form of f, which only allows non-convexity in $g_1 \circ S$ we obtain, similarly as above the following estimates:

$$\begin{aligned} |f'_u \delta u| &\le c(u)\|\delta u\| \\ c_0(u)|\delta u|^2 \le f''_u(\delta u, \delta u) &\le c_1(u)\|\delta u\|^2. \end{aligned}$$

3 Algorithmic framework

In this work we will follow the ideas of [16] and consider algorithms that are based on successive computation of low dimensional search spaces, minimization of a cubic model within these search spaces, and update of the model parameters. We consider the following conceptual algorithm for the step computation δx:

Algorithm 3.1. Start with initial guess x_0 and ω

repeat *(outer loop)*
  repeat *(inner loop)*
    Compute a non-trivial search space V *(cf. Section 3.2)*
    Compute a minimizer δx of $m^{\omega}_{x_k}$ in V *(cf. Section 3.1)*
    Update the parameter ω *(cf. Section 3.4)*
  until acceptance test satisfied *(cf. Section 3.3)*
  Update: $x_{k+1} = x_k + \delta x$
until convergence test satisfied

This general algorithm offers room for a large variety of implementations. They may differ in the way V is computed, ω is updated, and iterates are accepted. In particular, it includes the possibility to keep V, while only ω is updated.

In practical implementations, a convergence test (last line of the algorithm) may check, whether $\|f'_{x_{k+1}}\|$ is sufficiently small. For our theoretical purpose, where we consider a possibly infinite sequence of iterates, it is sufficient to check whether $f'_{x_{k+1}} = 0$.

3.1 Directional model minimizers

As a minimal requirement, we suppose that the trial corrections δx minimize m^{ω}_x along $\mathrm{span}\{\delta x\}$. We call such corrections *directional minimizers*. They are easy to compute and have nice properties.

Existence of a minimizer of m_x^ω may not hold due to a lack of convexity or compactness. However, if we pick finite dimensional subspaces, the minimizers within these subspaces, and in particular directional minimizers exist.

Theorem 3.2. *For a directional minimizer δx of m_x^ω it holds $f'_x\delta x \le 0$ and*

$$0 = f'_x\delta x + H_x(\delta x, \delta x) + \frac{\omega}{2}R_x(\delta x), \tag{11}$$

$$m_x^\omega(\delta x) = \frac{1}{2}f'_x\delta x - \frac{\omega}{12}R_x(\delta x) \tag{12}$$

$$= -\frac{1}{2}H_x(\delta x, \delta x) - \frac{\omega}{3}R_x(\delta x). \tag{13}$$

Proof. By the symmetry of the term $H_x(\delta x, \delta x) + \omega/6R_x(\delta x)$, it follows that $m_x^\omega(-\delta x) < m_x^\omega(\delta x)$ if $f'_x\delta x > 0$. Hence, a directional minimizer of m_x^ω satisfies $f'_x\delta x \le 0$.

As first order optimality conditions for a minimizer δx of m_x^ω we compute:

$$0 = (m_x^\omega)'(\delta x)v = f'_x v + H_x(\delta x, v) + \frac{\omega}{6}R'_x(\delta x)v \quad \forall v \in \operatorname{span}\{\delta x\}. \tag{14}$$

and thus, by homogeneity (4) of R_x we conclude $R'_x(\delta x)v = 3R_x(\delta x)v$ and thus (11). Inserting this into the definition of m_x^ω, we obtain (12) – (13). □

The following basic property is a simple consequence:

Lemma 3.3. *Let $\delta x(\omega)$ be the directional model minimizers along a fixed direction Δx for given ω. We have*

$$\lim_{\omega\to\infty} \delta x(\omega) = 0. \tag{15}$$

Proof. Fix $\omega_0 > 0$ and denote the corresponding directional minimizer in our direction by Δx. For any other $\omega > 0$ we have $\delta x(\omega) = \lambda\Delta x$ with $\lambda > 0$. Inserting this into (11) and dividing by λ we obtain the following quadratic equation for λ:

$$0 = f'_x\Delta x + \lambda H_x(\Delta x, \Delta x) + \lambda^2\frac{\omega}{2}R_x(\Delta x)$$

Since all coefficients of this quadratic polynomial, except for ω remain constant (15) follows from a straightforward computation. □

3.2 Acceptable search directions

Since we do not want to use degenerate directions for our directional minimizers, we impose a "fraction of Cauchy decrease" type condition. Classically this involves the explicit computation of a direction of steepest descent Δx^{SD} in each step of the outer loop. Its purpose is to establish a link between primal quantities δx and dual quantities f'_x. We emphasize that steepest descent directions depend on the choice of the norm $\|\cdot\|$.

In $\mathbb{R}^n$ the direction of steepest descent of f is commonly defined via the standard scalar product of $\mathbb{R}^n$ so that the negative gradient is given by $\Delta x^{SD} = -\nabla f(x) = -(f'_x)^T$. In the infinite dimensional setting this simple computation via transposition is not possible anymore (how would you "transpose" a function?), and has to be performed via solving the problem

$$\text{Find } \Delta x^{SD}: \quad f'_x\Delta x^{SD} = -\|f'_x\|\|\Delta x^{SD}\| \tag{16}$$

Usually, this is a non-trivial computation, depending on the choice of the norm. In the case of Hilbert spaces, it amounts in the computation of the Riesz isomorphism. Moreover, via the Cauchy step the choice of the norm $\|\cdot\|$ will directly influence our algorithm. This means that a good choice will improve the performance of our algorithm, while a a poor choice will degrade its performance.

In many cases the analytically straightforward choice of $\|\cdot\|$ will lead to a rather expensive computation of Δx^{SD} via (16). For example, if $\|\cdot\|_{H^1}$ is used, then Δx^{SD} has to be computed from f'_x via the solution of an elliptic partial differential equation. It is sufficient, however, to compute *quasi*-steepest descent directions, which satisfy:

Condition 3.4. *Let $1 \le \mu < 0$ be fixed. We compute at the beginning of each inner loop a* fixed quasi steepest descent direction Δx^C*, which satisfies*

$$f'_x \Delta x^C \le -\mu \|f'_x\| \|\Delta x^C\| \tag{17}$$

Within each inner loop, in which ω is adjusted, quasi Cauchy steps δx^C *are computed as directional minimizers of m_x^ω in direction of Δx^C.*

Often these steps are much cheaper to compute via a *preconditioner* than exact steepest descent directions. In our H^1-example this could be one cycle of a multigrid method.

Note that δx^C results from a *scaling* of Δx^C:

$$\delta x^C = \lambda(\omega)\Delta x^C, \quad \lambda > 1.$$

In our flexible framework R_x can be chosen quite independently of $\|\cdot\|$. This results in a modification of the classical Cauchy decrease condition. This modification penalizes irregular search directions, i.e., directions, where $\|\delta x\|^3 \gg R_x(\delta x)$ and thus avoids that iterates leave $(X, \|\cdot\|)$:

Condition 3.5. *Let $1 \ge \beta_0 > 0$ be fixed and δx^C be the quasi Cauchy step of m_x^ω. For δx define*

$$\beta := \beta_0 \max\left\{1, \left(\frac{R_x(\delta x^C)}{\|\delta x^C\|^3} \cdot \frac{\|\delta x\|^3}{R_x(\delta x)}\right)^{1/2}\right\}. \tag{18}$$

Then choose δx as a directional minimizer of m_x^ω, such that

$$m_x^\omega(\delta x) \le \beta m_x^\omega(\delta x^C). \tag{19}$$

The criterion (18) reduces to $\beta = \beta_0$, if either $\delta x = \delta x^C$, so that δx^C is acceptable, or $R_x(\cdot) = \|\cdot\|^3$, which is the standard case.

Lemma 3.6. *The following inequality holds for δx^C, as defined in Condition 3.5:*

$$\mu \|f'_x\| \|\delta x^C\| \le H_x(\delta x^C, \delta x^C) + \frac{\omega}{2} R_x(\delta x^C). \tag{20}$$

Let δx be a directional minimizer that satisfies (19)*. Then*

$$\frac{R_x(\delta x)}{\|\delta x\|} \ge \beta_0^2 \frac{R_x(\delta x^C)}{\|\delta x^C\|}. \tag{21}$$

Proof. From (11) and the fact that δx^C is a quasi steepest descent direction, we conclude

$$\mu \|f'_x\| \|\delta x^C\| \overset{(17)}{\le} |f'_x \delta x^C| = H_x(\delta x^C, \delta x^C) + \frac{\omega}{2} R_x(\delta x^C),$$

which implies (20).

To show (21) assume first that $f_x'\delta x + \beta\|f_x'\|\|\delta x^C\| \le 0$. Then

$$\|f_x'\|\|\delta x\| \ge |f_x'\delta x| = -f_x'\delta x \ge \beta\|f_x'\|\|\delta x^C\|,$$

and thus $\|\delta x\| \ge \beta\|\delta x^C\|$. Inserting (18) we obtain

$$\left(\frac{R_x(\delta x)}{\|\delta x\|}\right)^{1/2} \ge \beta_0 \left(\frac{R_x(\delta x^C)}{\|\delta x^C\|}\right)^{1/2}$$

which implies (21) in this case.

Otherwise, we use (12) for δx, (19), and (12) for δx^C to compute

$$\begin{aligned}\frac{\omega}{6}R_x(\delta x) &= f_x'\delta x - 2m_x^\omega(\delta x) \ge f_x'\delta x - 2\beta m_x^\omega(\delta x^C)\\ &= \underbrace{f_x'\delta x + \beta\|f_x'\|\|\delta x^C\|}_{\ge 0} + \beta\frac{\omega}{6}R_x(\delta x^C) \ge \beta\frac{\omega}{6}R_x(\delta x^C),\end{aligned}$$

and thus, $R_x(\delta x) \ge \beta R_x(\delta x^C)$. Inserting once again (18) we get

$$\left(\frac{R_x(\delta x)}{\|\delta x\|}\right)^{3/2} \ge \beta_0 \left(\frac{R_x(\delta x^C)}{\|\delta x^C\|}\right)^{3/2},$$

and thus by $\beta_0 \le 1$ also (21). □

Computing search directions. Let us briefly discuss known ways to compute acceptable search directions with the aim to explore a couple of possible alternatives. We restrict ourselves to giving a quick overview. A more detailed development is postponed to a forthcoming publication.

The simplest way to deal with optimization problems in function space numerically is to discretize them first, for example by finite elements. The idea of the *Ritz method* is to restrict the problem $\min_X f$ to a finite dimensional finite element subspace $X_h \subset X$ and solve $\min_{X_h} f$. It is, however, important to preserve the structure available from the infinite dimensional problem. In particular, the norms and thus the steepest descent directions used in the infinite dimensional context should be kept in the finite dimensional problem, if efficient behavior of our algorithm for fine direcretizations is the aim. We end up with a large finite dimensional problem, whose size grows, as the discretization parameter h tends to zero. Thus, for step computations we only consider iterative methods. Canonical candidates are Krylov-subspace methods, in particular preconditioned (truncated) cg or Lanczos methods (cf. e.g. [4, Chap. 5] and the references therein). If a reasonable *preconditioner* is available for these methods, then it can be used for the computation of a quasi Cauchy step, as we will explain in the following.

Let $(X_h, \|\cdot\|)$ be a finite dimensional subspace of $(X, \|\cdot\|)$ and consider a symmetric positive definite preconditioner $M^{-1} : X_h^* \to X_h$, where $M : X_h \to X_h^*$ satisfies the following (sometimes called "spectral") equivalence condition (with constants that, ideally, do not depend on h).

$$c_M\|v\|^2 \le (Mv)v \le \|M\|\|v\|^2.$$

Here $\|M\|$ is the operator norm of M, i.e., $\sup_{\|v\|=1}\|Mv\|$, for which it is well known that (by symmetry of M) it is the smallest constant for which the right inequality holds. Then $\Delta x^C := M^{-1}f_x'$ is a quasi steepest descent direction:

$$f_x'\Delta x^C = (M\Delta x^C)\Delta x^C \ge c_M\|\Delta x^C\|^2 \ge \frac{c_M}{\|M\|}\|f_x'\|\|\Delta x^C\|,$$

and thus can be used to compute quasi Cauchy steps in our algorithm. Moreover, if in addition (10) holds close to a local minimizer, then M is even spectrally equivalent to H_x, by assumption. Thus, locally we expect efficient behavior of our iterative solver. The construction of optimal preconditioners, in particular for H^1 elliptic problems is a well established and broad topic of research in numerical analysis and scientific computing (cf. e.g. [7, 1] and the references therein).

An alternative to solving a discretized optimization problem, i.e., choosing a fixed finite dimensional subspace X_h of X is to stay within the original infinite dimensional space X. Of course, also in this case the iterates and the trial corrections still have to be computed in finite dimensional subspaces, but these subspaces can be enlarged adaptively during the iteration, so that ultimately the sequence x_k may converge to a minimizer x_* of the original, infinite dimensional problem, and not just to a solution of the discretized problem. This means that techniques of adaptive grid refinement and optimization can be merged. We will not delve into this issue at this moment and leave it as a topic of future research to work out the details of this approach in the context of our proposed algorithmic framework. Works in this direction can be found, for example, in [17, 5].

3.3 Acceptance of trial corrections

After a directional minimizer of our model has been computed, and serves as a trial correction, we have to decide, whether this trial correction is acceptable as an optimization step. For this purpose we impose the following relative acceptance criterion, which is well known and popular in trust-region methods. To this end, let us define the ratio of decrease in f and in the model m_x^ω:

$$\eta_k := \frac{f(x+\delta x) - f(x)}{m_x^\omega(\delta x)} = \frac{f(x+\delta x) - f(x)}{m_x^\omega(\delta x) - m_x^\omega(0)}. \tag{22}$$

Recall that we have chosen m_x^ω in a way that $m_x^\omega(0) = 0$. Since $m_x^\omega(\delta x) < 0$, we see that $\eta_k > 0$ yields a decrease of f, and $\eta_k = 0$ means that f has remained constant. This yields the following condition:

Condition 3.7. *Choose $\underline{\eta} \in]0, 1[$. A trial correction δx is accepted, if it satisfies the condition*

$$\eta_k \geq \underline{\eta}. \tag{23}$$

Otherwise it is rejected.

3.4 Adaptive choice of ω

In this section we discuss the adaptive choice of ω in $m_x^\omega(\delta x)$. In [2] the choice of ω is made according to a classification of the steps into “unsuccessful”, “successful” and “very successful”. Here, in contrast, we compute ω via finite differences, which yields the following formula

$$\omega := \frac{6(f(x+\delta x) - q_x(\delta x))}{R_x(\delta x)} \overset{(2)}{\leq} \frac{6w_x(\delta x)}{R_x(\delta x)}, \tag{24}$$

and equip it with some save-guard restrictions. In order to guarantee positivity of ω and to avoid oscillatory behavior, we assume that the algorithm provides restrictions on updates

$\omega_{old} \to \omega_{new}$ to guarantee:

$$\omega_{new} > 0$$
$$\omega_{new} \le \frac{1}{\rho}\left(\omega_{old} + C_\omega + 2\frac{C_{f'}|f'_x\delta x| + |H_x(\delta x, \delta x)|}{R_x(\delta x)}\right) \qquad \text{for some } 0 < \rho < 1. \tag{25}$$

Positivity of ω is, of course, a basic requirement which guarantees that the term R_x is present throughout the computation. The second condition (25) inhibits that ω is increased too quickly, with the result that the next trial correction has to be chosen much shorter than the previous one. However, in a certain range (corresponding to C_ω) the increase can be performed freely. If R_x is much smaller than the remaining terms of m_x^ω a fast increase of ω is also possible. Technically, this restriction enters into the global convergence proof in (50), below.

The following theory will cover algorithms, that respect these restrictions, and increase ω after a rejected trial correction, according to

$$\text{After rejected trial correction:} \qquad \omega_{new} \ge \min\{\omega, C_+\omega_{old}\} \qquad C_+ > 1. \tag{26}$$

Any algorithm that does not allow this, is likely to forbid a useful increase of ω and get stuck in an inner loop. This restriction and (25) do not interfere, if we take, e.g., $\rho \le C_+^{-1}$. Technically this condition is used at the beginning of the proof of Theorem 3.8.

To obtain fast local convergence under most general assumptions we do not increase ω if η_k is very close to 1. Let us chose $\overline{\eta} \in [\underline{\eta}, 1[$ and state

$$\text{If } \ \eta_k > \overline{\eta} \ \text{ then} \qquad \omega_{new} \le \omega_{old}. \tag{27}$$

Finally, we impose the restriction on our algorithm that after an increase of ω w_x is essentially estimated from below by ωR_x:

$$\text{If} \quad \omega_{new} \ge \omega_{old}: \qquad \omega_{new}R_x(\delta x) \le C_w w_x(\delta x). \tag{28}$$

By (24) this is easy to realize, if we impose e.g., $\omega_{new} \le \max\{\omega, \omega_{old}\}$.

From a practical point of view, we have to take into account that our third order estimate is subject to round-off errors and can be replaced by a lower order estimate, which involves first derivatives at slightly higher cost [16] .

In our framework we deliberately dispense with a-priori restrictions like $0 < \underline{\omega} \le \omega$. Such restrictions can, and should of course, be added in finite precision arithmetic.

3.5 Finite termination of inner loops

Next, we show that each inner loop of our algorithm accepts a finite ω after finitely many updates and thus terminates finitely. Hence, in the following we consider fixed x and a sequence ω_i of parameters and δx_i of trial corrections, computed by our algorithm.

Theorem 3.8. *Assume that $\|f'_x\| \ne 0$ and let $1 > \underline{\eta} > 0$ and $\beta > 0$. Moreover, assume that the lower bounds of (5) and (6) hold. Further, assume that (7) holds for fixed x.*

Then:

(i) If a trial correction is rejected, then ω is increased afterwards.

(ii) The inner loop terminates successfully after finitely many iterations.

Proof. Let us consider a single inner loop. Our aim is to exclude that infinitely many trial iterates δx_i are rejected within this inner loops. To this end we will first show that in such a case $\omega_i \to \infty$ and then derive a contradiction.

First, we show that if (23) is violated, then $\omega_{new} \geq \min\{C_+, 3/2 - \underline{\eta}/2\}\omega_{old}$. Assume that $\omega_{new} < C_+\omega_{old}$. Then violation of (23) implies

$$\begin{aligned}
\omega_{new} \geq \omega &= \frac{6}{R_x(\delta x)}(f(x+\delta x) - q_x(\delta x)) \\
&= \frac{6}{R_x(\delta x)}\left(f(x+\delta x) - f(x) - m_x^{\omega_{old}}(\delta x) + \frac{\omega_{old}}{6}R_x(\delta x)\right) \\
&> \frac{6}{R_x(\delta x)}(\underline{\eta} - 1)m_x^{\omega_{old}}(\delta x) + \omega_{old} \\
&= \frac{6}{R_x(\delta x)}(1-\underline{\eta})\left(-1/2 f_x'\delta x + \frac{\omega_{old}}{12}R_x(\delta x)\right) + \omega_{old} \\
&\geq ((1-\underline{\eta})/2 + 1)\omega_{old}.
\end{aligned}$$

Thus, ω is increased at least by a fixed factor above 1. Next, assume for contradiction that (23) fails infinitely often during successive updates from ω_{old} to ω_{new}. Thus, there is a sequence $\omega_i \to \infty$, corresponding Cauchy steps δx_i^C, and trial corrections δx_i. Lemma 3.3 yields $\delta x_i^C = \lambda(\omega_i)\Delta x^C \to 0$ and thus

$$\frac{\omega_i R_x(\delta x_i^C)}{2\|\delta x_i^C\|} \overset{(20)}{\leq} \mu\|f_x'\| - \underbrace{\frac{H_x(\lambda\Delta x^C, \lambda\Delta x^C)}{\|\lambda\Delta x^C\|}}_{\to 0} \to \mu\|f_x'\| > 0.$$

since $\|f_x'\| \neq 0$. It thus follows from (21) that there exists a constant $W_0 > 0$ such that

$$\frac{\omega_i R_x(\delta x_i)}{\|\delta x_i\|} \geq W_0 > 0. \tag{29}$$

Since ω is increased at every step, it follows from our restriction (28) that $\omega_{i+1}R_x(\delta x_i)$ is an estimate from below of $C_w w_x(\delta x_i)$. Thus, we conclude that

$$0 < W_0 \leq \frac{\omega_i R_x(\delta x_i)}{\|\delta x_i\|} \leq \frac{\omega_{i+1} R_x(\delta x_i)}{\|\delta x_i\|} \leq \frac{C_w w_x(\delta x_i)}{\|\delta x_i\|},$$

and hence by our differentiability assumption (8) from Lemma 2.1 there is $R_0 > 0$ such that

$$\omega_i R_x(\delta x_i) \geq R_0. \tag{30}$$

With this, we compute from (11) (using $f_x'\delta x_i = -|f_x'\delta x_i|$)

$$\frac{|f_x'\delta x_i|}{\omega_i R_x(\delta x_i)} = \frac{H_x(\delta x_i, \delta x_i)}{\omega_i R_x(\delta x_i)} + \frac{1}{2}.$$

By (5), if the middle term including the hessian has a negative contribution it vanishes asymptotically:

$$\lim_{k\to\infty}\left|\frac{\min\{H_x(\delta x_i, \delta x_i), 0\}}{\omega_i R_x(\delta x_i)}\right| \overset{(5)}{\leq} \lim_{k\to\infty}\frac{|\gamma|\|\delta x_i|^2}{\omega_i R_x(\delta x_i)} \overset{(6)}{\leq} \lim_{k\to\infty}\frac{|\gamma|}{\omega_i^{2/3}(\omega_i R_x(\delta x_i))^{1/3}} \overset{(30)}{=} 0.$$

Thus we conclude that there is an index k_0 and a constant $M_0 > 0$ such that

$$\frac{|f_x'\delta x_i|}{\omega_i R_x(\delta x_i)} \geq M_0 > 0 \qquad \forall k \geq k_0 \tag{31}$$

and thus via (29) that also

$$\liminf_{k\to\infty} \frac{|f_x'\delta x_i|}{\|\delta x_i\|} > 0. \tag{32}$$

However, as a consequence of (31), (30), and (6) we have for $k \geq k_0$:

$$\frac{\|\delta x_i\|}{|\delta x_i|} \geq \frac{|f_x'\delta x_i|}{|\delta x_i|\|f_x'\|} \overset{(31)}{\geq} M_0 \frac{\omega_i R_{x_i}(\delta x_i)}{|\delta x_i|\|f_x'\|} \overset{(6)}{\geq} M_0 \frac{\omega_i R_x(\delta x_i)^{2/3}}{\|f_x'\|} \geq \frac{M_0 R_0^{2/3}}{\|f_x'\|}\omega_i^{1/3} \to \infty.$$

Thus, $|\delta x_i|/\|\delta x_i\| \to 0$ and by Lemma 2.2 we conclude weak convergence $\delta x_i/\|\delta x_i\| \rightharpoonup 0$ in $(X, \|\cdot\|)$. This, however, implies

$$\lim_{k\to\infty} \frac{|f_x'\delta x_i|}{\|\delta x_i\|} = 0.$$

in contradiction to (32). $\square$

4 Convergence Theory

In this section we will establish first order global convergence, and second order local convergence results. In the following we will consider a sequence x_k, generated by our algorithm, corresponding derivatives f'_{x_k}, and accepted corrections δx_k with parameters ω_k. In the whole section we exclude the trivial case that $f'_{x_k} = 0$, for some k, which leads to finite termination of our algorithm. Moreover, we may assume that the sequence $f(x_k)$ is bounded from below. Otherwise our algorithm fulfills its purpose by generating a sequence $f(x_k) \to -\infty$.

4.1 Global Convergence

Under mild assumptions we will show that our algorithm cannot converge to non-stationary points, while slightly stronger assumptions yield $\|f'_{x_k}\| \to 0$. Our technique will be to derive a contradiction to the case that $\|f'_{x_k}\|$ remains bounded away from zero. Taking this into account, we see that our theory will still work, if all algorithmic constants $(\underline{\eta}, \overline{\eta}, \mu, \beta_0, \dots)$ depend on $\|f_x'\|$ and degenerate only if $\|f_x'\| \to 0$.

The first two lemmas that we will prove are based on the acceptance criteria (23) (actual vs. predicted reduction) and (19) (modified fraction of Cauchy decrease) only. Recall that δx_k^C are quasi Cauchy steps, defined in Condition 3.4.

Lemma 4.1. *Assume that the sequence $f(x_k)$ is bounded from below. Assume that the successful steps δx_k are chosen as directional minimizers, such that* (23) *and* (19) *hold. Then*

$$\sum_{k=0}^{\infty} \|f'_{x_k}\|\|\delta x_k^C\| < \infty \tag{33}$$

$$\sum_{k=0}^{\infty} \omega_k R_{x_k}(\delta x_k^C) < \infty, \qquad \sum_{k=0}^{\infty} \omega_k R_{x_k}(\delta x_k) < \infty. \tag{34}$$

Proof. We use (23) and (19) to compute

$$\begin{aligned}
f(x_{k+1}) - f(x_k) &\overset{(23)}{\leq} \underline{\eta} m_{x_k}^{\omega_k}(\delta x_k) \overset{(11)}{=} \underline{\eta}\left(\frac{1}{2} f'_{x_k}\delta x_k - \frac{\omega_k}{12} R_{x_k}(\delta x_k)\right) \\
&\overset{(19)}{\leq} \underline{\eta}\beta_0 m_{x_k}^{\omega_k}(\delta x_k^C) \overset{(11)}{=} \underline{\eta}\beta_0 \left(\frac{1}{2} f'_{x_k}\delta x_k^C - \frac{\omega_k}{12} R_{x_k}(\delta x_k^C)\right) \\
&\leq \frac{\underline{\eta}\beta_0}{2} f'_{x_k}\delta x_k^C \overset{(17)}{\leq} -\frac{\mu\underline{\eta}\beta_0}{2}\|f'_{x_k}\|\|\delta x_k^C\| \leq 0.
\end{aligned}$$

By monotonicity and boundedness

$$\sum_{k=0}^{\infty} f(x_{k+1}) - f(x_k) = \underline{f} - f(x_0) > -\infty,$$

and by our chain of inequalities we conclude (33) and (34). □

The main observation here is that $\|f'_{x_k}\|\|\delta x_k^C\| \to 0$, and it remains to prevent $\|\delta x_k^C\|$ from becoming too small, compared to $\|f'_{x_k}\|$ in order to force $\|f'_{x_k}\|$ to become small. The extraordinary role of δx_k^C has its origin in the acceptance criterion (19), which compares all steps to the Cauchy steps.

To obtain a quick understanding of the situation, take a look at (20) and observe the following relation:

$$\mu\|f'_{x_k}\| \overset{(20)}{\leq} \frac{H_{x_k}(\delta x_k^C, \delta x_k^C)}{\|\delta x_k^C\|} + \frac{\omega_k}{2}\frac{R_{x_k}(\delta x_k^C)}{\|\delta x_k^C\|}.$$

We would like to exclude that case that $\|f'_{x_k}\|$ is bounded away from zero, which in turn implies $\|\delta x_k^C\| \to 0$ by (33). Taking into account the upper bounds (5) for H_x and (6) for R_x, we see that

$$\lim_{k\to\infty} \frac{H_{x_k}(\delta x_k^C, \delta x_k^C)}{\|\delta x_k^C\|} = 0, \tag{35}$$

$$\lim_{k\to\infty} \frac{R_{x_k}(\delta x_k^C)}{\|\delta x_k^C\|} = 0. \tag{36}$$

In fact for all the following results we may replace (5) and (6) by these weaker results. This might be a useful generalization in the case where regularity theory yields a-priori bounds on δx^C in a stronger norm than $\|\cdot\|$. In the context of partial differential equations additional regularity results are frequently encountered (cf. e.g. [8]). In this case we may gain additional flexibility in the choice of R_x.

In view of (35) and (36), which exclude that our iteration is choked off by H_x and R_x being overly large, the "bad case" can only happen, if ω_k is increased too rapidly. Under smoothness assumptions on f that imply boundedness of ω_k (a global Lipschitz condition on f'') we would be finished at this point. To cover the more general case (see (42) and (43), below), we have to invest some more theoretical work. Let us start with collecting some simple consequences of $\|f'_{x_k}\|$ being bounded away from 0:

Lemma 4.2. *Suppose that x_k is a sequence, such that $f(x_k)$ is bounded from below. For fixed $\nu > 0$ define the set*

$$\mathcal{L}^\nu := \{k : \|f'_{x_k}\| \geq \nu\}.$$

Assume that (35) and (36) hold for δx_k^C along the sequence of iterates x_k for $k \in \mathcal{L}^\nu$.

For iteration k let δx_k be the accepted step, and δv_k any directional minimizer of $m_{x_k}^{\omega_k}$ that satisfies (19)*. Then*

$$\inf_{k\in\mathcal{L}^\nu} \frac{\omega_k R_x(\delta v_k)}{\|f'_{x_k}\|\|\delta v_k\|} > 0, \tag{37}$$

$$\sum_{k\in\mathcal{L}^\nu} \|\delta x_k\| < \infty, \tag{38}$$

$$\lim_{k\in\mathcal{L}^\nu\to\infty} \omega_k = \infty. \tag{39}$$

Proof. The assertions are trivial or void, if $\mathcal{L}^\nu$ is finite, so it remains to consider the infinite case.

From $\|f'_{x_k}\| \ge \nu$ we get $\|\delta x_k^C\| \to 0$ due to (33) and thus, via (20) and (35):

$$\lim_{k\to\infty} \frac{\omega_k R_{x_k}(\delta x_k^C)}{2\|f'_{x_k}\|\|\delta x_k^C\|} \overset{(20)}{\ge} \lim_{k\to\infty} \Big(\mu - \underbrace{\frac{H_{x_k}(\delta x_k^C, \delta x_k^C)}{\|f'_{x_k}\|\|\delta x_k^C\|}}_{\overset{(35)}{\to}0}\Big) = \mu. \tag{40}$$

By (21) we also have

$$\liminf_{k\to\infty} \frac{\omega_k R_{x_k}(\delta v_k)}{\|f'_{x_k}\|\|\delta v_k\|} \ge 2\beta_0^2 > 0$$

and thus (37). Now (38) follows from (34) and (37) via the computation

$$\begin{aligned}\sum_{k\in\mathcal{L}^\nu} \|f'_{x_k}\|\|\delta x_k\| &= \sum_{k\in\mathcal{L}^\nu} \omega_k R_{x_k}(\delta x_k) \left(\frac{\omega_k R_{x_k}(\delta x_k)}{\|f'_{x_k}\|\|\delta x_k\|}\right)^{-1} \\ &\le \sum_{k\in\mathcal{L}^\nu} \omega_k R_{x_k}(\delta x_k) \left(\inf_{k\in\mathcal{L}^\nu} \frac{\omega_k R_{x_k}(\delta x_k)}{\|f'_{x_k}\|\|\delta x_k\|}\right)^{-1} < \infty,\end{aligned}$$

and the fact that $\|f'_{x_k}\|$ is bounded away from 0. By (40) we compute, using $\|\delta x_k^C\| \to 0$ and (36):

$$2\|f'_{x_k}\|\omega_k^{-1} \overset{(40)}{\le} c\frac{R_{x_k}(\delta x_k^C)}{\|\delta x_k^C\|} \overset{(36)}{\to} 0.$$

From that (39) follows by the fact that $\|f'_{x_k}\|$ is bounded away from 0. □

An important conclusion of this lemma is that if $\mathcal{L}^\nu = \mathbb{N}$, then by (38) x_k is a Cauchy sequence in $(X, \|\cdot\|)$, and thus x_k converges to some limit x_*. Thus, to exclude this case in a proof by contradiction we may always assume that $x_k \to x_*$.

Up to now, the smoothness of f and the lower bound in the Gårding inequality (5) did not enter our considerations. In the next lemma, which is the main step of our study, we will take this and the save-guard restrictions (28) and (25) on the update of ω into account. We will perform this step under the most general assumptions.

We are interested in the case that our algorithm converges to a non-stationary point. We show in this case that the set

$$\mathcal{I} := \{k \in \mathbb{N} : \text{ the inner loop at } x_k \text{ computes at least one rejected trial correction }\}$$

is infinite. Moreover, we will see that in this case either f is not smooth enough, or H_x is not regular enough. Two cases, which are excluded by our set of assumptions in Section 2.

Lemma 4.3. *Suppose that $x_k \to x_*$ in $(X, \|\cdot\|)$, such that $f(x_k)$ is bounded from below and $f'_{x_k} \to f'_* \neq 0$ in $(X, \|\cdot\|)^*$. Assume further that along x_k (35) and (36) hold for δx_k^C and the smoothness relation (7) holds for δx_k.*

Then $\mathcal{I}$ is infinite, and for the sequence of the last rejected trial corrections δx_k^l in each inner loop for $k \in \mathcal{I}$ there exists a constant $W_0 > 0$, such that:

$$\frac{C_w w_{x_k}(\delta x_k^l)}{\|\delta x_k^l\|} \geq \frac{\omega_k R_{x_k}(\delta x_k^l)}{\|\delta x_k^l\|} \geq W_0 > 0, \tag{41}$$

Additionally, for any infinite subset of $\mathcal{I}$ at least one of the following conditions is ***violated:***

(i) Smoothness: (8) holds for δx_k^l, along x_k, i.e.,

$$(41) \qquad \textit{implies} \qquad \exists R_0 > 0 : \omega_k R_{x_k}(\delta x_k^l) \geq R_0. \tag{42}$$

(ii) Boundedness of hessians from below along rejected directions:

$$\exists \gamma^l > -\infty : H_{x_k}(\delta x_k^l, \delta x_k^l) \geq \gamma^l |\delta x_k^l|^2. \tag{43}$$

Proof. By our assumptions the sequence $f(x_k)$ is bounded from below and the sequence $\|f'_{x_k}\|$ converges to a non-zero value and thus is bounded above and bounded away from 0. Then by (35) Lemma 4.2 holds for $\mathcal{L}^\nu = \mathbb{N}$ for some $\nu > 0$, and in particular ω_k is increased infinitely many times due to (39). An increase of ω can occur in two cases: first, after an accepted trial correction δx_k, second after a rejected trial correction. In the first case, i.e., $\omega_{k+1} \geq \omega_k$, we compute by (28):

$$\omega_k R_{x_k}(\delta x_k) \leq \omega_{k+1} R_{x_k}(\delta x_k) \overset{(28)}{\leq} C_w w_{x_k}(\delta x_k)$$

and thus, by (37) we conclude:

$$\frac{C_w w_{x_k}(\delta x_k)}{\|f'_{x_k}\|\|\delta x_k\|} \overset{(28)}{\geq} \frac{\omega_k R_{x_k}(\delta x_k)}{\|f'_{x_k}\|\|\delta x_k\|} \overset{(37)}{\geq} c.$$

It follows that there is a constant $\tilde{W}_0 > 0$ independent of k, such that

$$\frac{w_{x_k}(\delta x_k)}{\|\delta x_k\|} \geq \tilde{W}_0 > 0 \tag{44}$$

for every k after which ω_k was increased. However, since $\|\delta x_k\| + \omega_k R_{x_k}(\delta x_k) \to 0$ by (38) and (34), (7) implies that the left hand side of (44) tends to 0 along the sequence $x_k \to x_*$. Thus, the *first case* can only happen finitely many times.

Hence, the *second case* must occur infinitely many times, i.e., there must be infinitely many rejected trial corrections, and thus infinitely many loops in which a trial correction is rejected. This means that $\mathcal{I}$ is a set of infinite size. We will assume that (i) and (ii) hold, at least for an infinite subset $\mathcal{J}$ of $\mathcal{I}$, and derive a contradiction. We divide the remaining argumentation into 3 steps. In the following we will consider $k \in \mathcal{J}$ only.

Step 1: For the inner loop k at x_k consider the *last rejected correction* δx_k^l with corresponding regularization parameter ω_k^l. Recall that δx_k^l, like every trial correction, is a directional minimizer of $m_{x_k}^{\omega_k^l}$. After rejection of δx_k^l the next regularization parameter corresponds to the final accepted trial correction in this loop δx_k and is thus denoted by ω_k.

Let $\delta x_k^{C,l}$ be the Cauchy step for ω_k^l and δx_k^C the Cauchy step for ω_k. Since by Theorem 3.8 $\omega_k \ge \omega_k^l$, we have

$$\delta x_k^{C,l} = \lambda \delta x_k^C \quad \text{with} \quad \lambda \ge 1.$$

Then by (21) and (37) for $\delta v_k = \delta x_k^C$, taking into account the boundedness properties of $\|f'_{x_k}\|$, we get a constant $c > 0$, such that

$$\begin{aligned} \frac{\omega_k R_{x_k}(\delta x_k^l)}{\|\delta x_k^l\|} &\overset{(21)}{\ge} \beta_0^2 \frac{\omega_k R_{x_k}(\delta x_k^{C,l})}{\|\delta x_k^{C,l}\|} = \frac{\omega_k R_{x_k}(\lambda \delta x_k^C)}{\|\lambda \delta x_k^C\|} \\ &= \frac{\omega_k \lambda^3 R_{x_k}(\delta x_k^C)}{\lambda \|\delta x_k^C\|} \ge \beta_0^2 \frac{\omega_k R_{x_k}(\delta x_k^C)}{\|\delta x_k^C\|} \overset{(37)}{\ge} c\|f'_{x_k}\| \ge c\nu > 0. \end{aligned} \tag{45}$$

By (28) this implies that there is a constant $W_0 > 0$, such that

$$\frac{C_w w_{x_k}(\delta x_k^l)}{\|\delta x_k^l\|} \ge \frac{\omega_k R_{x_k}(\delta x_k^l)}{\|\delta x_k^l\|} \ge W_0 > 0, \tag{46}$$

This and (46) in turn implies by (42) (cf. also (8) in Lemma 2.1) that there is a constant $R_0 > 0$, such that for these *rejected* trial corrections

$$\omega_k R_{x_k}(\delta x_k^l) \ge R_0 > 0, \tag{47}$$

in contrast to (34), which holds for *accepted* trial corrections.

Step 2: Now we are in the position to show that there are k_0 and M_0 such that

$$\frac{|f'_{x_k} \delta x_k^l|}{\omega_k R_{x_k}(\delta x_k^l)} \ge M_0 > 0 \qquad \forall k \in \mathcal{J} : \ k \ge k_0. \tag{48}$$

Here our save-guard restriction for the update $\omega_k^l \to \omega_k$ (25) comes into play, which reads now:

$$\frac{\rho}{2} \omega_k R_x(\delta x_k^l) \le \frac{\omega_k^l + C_\omega}{2} R_x(\delta x_k^l) + C_{f'_x} |f'_x \delta x_k^l| + |H_x(\delta x_k^l, \delta x_k^l)|. \tag{49}$$

We insert this relation into (11), the optimality condition for directional minimizers:

$$\begin{aligned} |f'_{x_k} \delta x_k^l| = -f'_{x_k} \delta x_k^l &\overset{(11)}{=} \frac{\omega_k^l}{2} R_{x_k}(\delta x_k^l) + H_{x_k}(\delta x_k^l, \delta x_k^l) \\ &\overset{(49)}{\ge} \frac{\rho \omega_k - C_\omega}{2} R_{x_k}(\delta x_k^l) - C_{f'} |f'_x \delta x_k^l| - |H_{x_k}(\delta x_k^l, \delta x_k^l)| + H_{x_k}(\delta x_k^l, \delta x_k^l) \\ &= \frac{\rho \omega_k - C_\omega}{2} R_{x_k}(\delta x_k^l) - C_{f'} |f'_x \delta x_k^l| + 2 \min\{H_{x_k}(\delta x_k^l, \delta x_k^l), 0\}, \end{aligned}$$

so that we obtain:

$$\frac{(1 + C_{f'})|f'_{x_k} \delta x_k^l|}{\omega_k R_{x_k}(\delta x_k^l)} \ge \frac{\rho}{2} - \frac{C_\omega}{2\omega_k} - 2 \left| \frac{\min\{H_{x_k}(\delta x_k^l, \delta x_k^l), 0\}}{\omega_k R_{x_k}(\delta x_k^l)} \right|. \tag{50}$$

Since $\omega_k \to \infty$, the second term on the right hand side of (50) tends to zero. Moreover, by (43), (6), and (47) the same is true for the third term:

$$\begin{aligned} \lim_{k \in \mathcal{J} \to \infty} \left| \frac{\min\{H_{x_k}(\delta x_k^l, \delta x_k^l), 0\}}{\omega_k R_{x_k}(\delta x_k^l)} \right| &\overset{(43)}{\le} \lim_{k \in \mathcal{J} \to \infty} \frac{|\gamma^l| \|\delta x_k^l\|^2}{\omega_k R_{x_k}(\delta x_k^l)} \overset{(6)}{\le} \lim_{k \in \mathcal{J} \to \infty} \frac{|\gamma^l|}{\omega_k R_{x_k}(\delta x_k^l)^{1/3}} \\ &\overset{(47)}{\le} \lim_{k \in \mathcal{J} \to \infty} \frac{|\gamma^l|}{\omega_k^{2/3} R_0^{1/3}} \overset{(39)}{=} 0. \end{aligned}$$

Hence, in the limit the left hand side of (50) is strictly positive, which implies (48).

Step 3: Let us finally derive our contradiction. Multiplication of (48) with the middle term in (46) implies on the one hand

$$\frac{|f'_{x_k}\delta x_k^l|}{\|\delta x_k^l\|} \geq M_0 W_0 > 0 \qquad \forall k \in \mathcal{J} : k \geq k_0. \tag{51}$$

On the other hand due to (48), (6), and (47) we have for $k \in \mathcal{J}, k \geq k_0$:

$$\frac{\|\delta x_k^l\|}{|\delta x_k^l|} \geq \frac{|f'_{x_k}\delta x_k^l|}{\|f'_{x_k}\||\delta x_k^l|} \overset{(48)}{\geq} M_0 \frac{\omega_k R_{x_k}(\delta x_k^l)}{\|f'_{x_k}\||\delta x_k^l|} \overset{(6)}{\geq} \left(\frac{M_0}{\|f'_{x_k}\|}\underbrace{(\omega_k R_{x_k}(\delta x_k^l))^{2/3}}_{\geq R_0 \text{ by } (47)}\right)\underbrace{\omega_k^{1/3}}_{\to\infty} \to \infty,$$

and thus

$$\lim_{k\in\mathcal{J}\to\infty} \frac{|\delta x_k^l|}{\|\delta x_k^l\|} = 0.$$

Via Lemma 2.2 we conclude that the normalized sequence $\delta x_k^l/\|\delta x_k^l\|$ converges to 0 weakly in X. Since $f'_{x_k} \to f'_*$ strongly, we obtain a contradiction to (51):

$$\lim_{k\in\mathcal{J}\to\infty} \frac{|f'_{x_k}\delta x_k^l|}{\|\delta x_k^l\|} = f'_* 0 = 0.$$

This is due to a standard result in functional analysis which states that the duality product is continuous with respect to strong convergence in the dual space and weak convergence in the primal space. □

The following conclusions are more or less direct consequences of the previous lemma. For ease of reading we use slightly stronger, but more concise assumptions.

Theorem 4.4. *Let $x_* \in X$. Assume that f is Fréchet differentiable in a neighborhood of x_* and f' is continuous at x_*. Further, assume that $x_k \to x_*$ in $(X, \|\cdot\|)$ and that the Gårding inequality* (5) *holds along x_k. Moreover, assume that R_x satisfies* (6).

Then

$$f'_{x_*} = 0.$$

Proof. Since $x_k \to x_*$, also $f'_{x_k} \to f'_{x_*}$ in X^*. Moreover, (35) and (43) hold by (5). If $f'_{x_*} \neq 0$, then (41) and (42) contradict (7). □

If x_k does not converge, we can still show convergence properties for f', following the standard pattern that continuity of f' yields subsequential convergence of f' to 0, while uniform continuity yields convergence of the whole sequence.

Theorem 4.5. *Let f be Fréchet differentiable. Assume that $f(x_k)$ is bounded from below and H_x satisfies the Gårding inequality* (5). *Further, assume that R_x satisfies* (6).

(i) If f' is continuous, then

$$\liminf_{k\to\infty} \|f'_{x_k}\| = 0. \tag{52}$$

(ii) If f' is uniformly continuous, then

$$\lim_{k\to\infty} \|f'_{x_k}\| = 0.$$

Proof. For the purpose of contradiction we assume that $\|f'_{x_k}\|$ is bounded away from zero. Then x_k is a Cauchy sequence in X by (38), and hence convergent to a limit point x_* by completeness of X.

Under the assumptions of (i), this implies $\|f'_{x_k}\| \to 0$ by Theorem 4.4. This is a contradiction our premise that $\|f'_{x_k}\|$ is bounded away from zero, and hence (52) must hold.

It remains to assert that $\mathcal{L}^\nu$ is finite under the assumptions of (ii) for any $\nu > 0$. For this we use a standard trick (cf. e.g. [4, Thm 6.4.6]), exploiting uniform continuity of the function $x \to f'_x$. For any index $i \in \mathcal{L}^\nu$ choose the first index $k(i) \in \mathbb{N} \setminus \mathcal{L}^{\nu/2}$ that satisfies $k(i) > i$. Then $\{j : i \le j < k(i)\} \subset \mathcal{L}^{\nu/2}$ and by (38)

$$\lim_{i\to\infty} \|x_i - x_{k(i)}\| \le \lim_{i\to\infty} \sum_{j=i}^{j<k(i)} \|\delta x_j\| = 0$$

and thus, if $\mathcal{L}^\nu$ was infinite, $\|f'_{x_i} - f'_{x_{k(i)}}\| \to 0$. However, eventually

$$\|f'_{x_i}\| \ge \nu,$$
$$\|f'_{x_{k(i)}}\| \le \nu/2.$$

Hence, we have a contradiction and $\mathcal{L}^\nu$ must be finite. This argument holds for every $\nu > 0$ and thus implies $\lim_{k\to\infty} \|f'_{x_k}\| = 0$. □

4.2 Local convergence

Next we consider local convergence of our method towards a local minimizer. We will show in this context that damping factors λ approach 1 close to a local minimizer under additional assumptions. Whether this finally leads to local superlinear convergence or not depends mostly on the actual computed search directions δx_k. It it well known that Newton directions and certain quasi-Newton directions lead to fast local convergence. Here we will content ourselves with showing that our globalization scheme does not interfere with any method to compute search directions.

Let us start with some auxiliary estimates, which capture the effect of positive curvature of H_x along a directional minimizer. Theses estimates do not rely on a fraction of Cauchy decrease condition:

Lemma 4.6. *Let δx be a directional minimizer and $H_x(\delta x, \delta x) = \gamma\|\delta x\|^2$ with $\gamma \ge 0$. Then we have the following estimates:*

$$m_x^\omega(\delta x) \le -\frac{\gamma}{2}\|\delta x\|^2 \tag{53}$$

$$\gamma\|\delta x\| \le \|f'_x\|. \tag{54}$$

Proof. Equation (53) directly follows from (13), taking into account positivity of R_x.

Equation (11) yields

$$\gamma\|\delta x\|^2 \le \gamma\|\delta x\|^2 + \frac{\omega}{2}R_x(\delta x) = H_x(\delta x, \delta x) + \frac{\omega}{2}R_x(\delta x) = -f'_x\delta x \le \|f'_x\|\|\delta x\|$$

and thus (54). □

Our basic theoretical framework comprises the following assumptions, which we impose throughout the whole section. For fast local convergence we will later impose further smoothness assumptions.

Assumption 4.7. Let $x_* \in X$ be a local minimizer, and assume that there exists a neighborhood U of x_* with the following properties:

(i) The assumptions of Theorem 4.5(i) on global convergence hold in U.

(ii) For $\varepsilon > 0$ define the local level sets

$$L_\varepsilon := \{x \in U : f(x) \le f(x_*) + \varepsilon\} \subset U.$$

Assume that these sets form a neighborhood base of x_*, i.e., each neighborhood of x_* contains one of these level sets (and hence all with smaller ε). This implies that x_* is a local minimizer. The converse is not true, in general.

(iii) We have the estimate

$$\exists \alpha < \infty \ : \ f(x) - f(x_*) \le \alpha \|f'_x\| \|x - x_*\| \quad \forall x \in U.$$

This holds with $\alpha = 1$, if f is convex in U, and implies, together with (ii) that x_* is an isolated critical point.

(iv) The ellipticity assumption (10) for H_x holds in U:

$$\exists \gamma > 0 : \quad \gamma \|\delta x\|^2 \le H_x(\delta x, \delta x).$$

If f is twice differentiable and $H_x = f''_x$, then this implies convexity of f in U and thus (iii).

It follows from continuity of f that the interior of L_ε is non-empty, and (ii) implies via differentiability of f that $f'_{x_*} = 0$. Alternatively to (iii) we could assume continuous invertibility of the mapping $x \to f'_x$.

First we show that if our algorithm comes close to a local minimizer with the above properties, then it will converge towards this minimizer.

Lemma 4.8. *If Assumption 4.7 holds, then there exists $\varepsilon_0 > 0$ such that if $x \in L_\varepsilon$, and δx is an acceptable directional minimizer then $x + \delta x \in L_\varepsilon$ for all $0 < \varepsilon < \varepsilon_0$.*

Proof. By Assumption 4.7(ii) we can choose for any neighborhood $V \subset U$ of x_* an $\varepsilon > 0$, such that $L_\varepsilon \subset V$. Recall that H_x is uniformly elliptic on U and thus on V with a constant $\gamma > 0$. By continuity of f'_x we can in turn choose V, such that $\|f'_x\| \le \gamma^{-1}\nu$ for every $x \in V$, for every given $\nu > 0$. It follows by (54) that $\|\delta x\| \le \nu$ for every acceptable directional minimizer, and thus $x + \delta x \in U$, as long as V and ν have been chosen sufficiently small, and $x \in L_\varepsilon \subset V$. Thus, we conclude by the descent property that $x + \delta x \in L_\varepsilon \subset V$, again. □

Theorem 4.9. *Suppose that Assumption 4.7 holds. If the sequence of iterates, generated by our algorithm comes sufficiently close to x_*, then it converges to x_*.*

Proof. By Lemma 4.8 the sequence, generated by our algorithm remains in L_ε, as long as one iterate comes sufficiently close to x_*. Thus, $\|x_k - x_*\|$ remains bounded. Theorem 4.5 implies $\|f'_{x_{k_i}}\| \to 0$, at least for a subsequence x_{k_i}, and thus

$$f(x_{k_i}) - f(x_*) \le \alpha \|f'_{x_{k_i}}\| \|x_{k_i} - x_*\| \to 0.$$

So, for each $\varepsilon > 0$, $x_{k_i} \in L_\varepsilon$, eventually. Since x_k does not leave level sets by Lemma 4.8, the same holds for the whole sequence. Since the level sets form a neighborhood base of x_*, we conclude that $x_k \to x_*$. □

Finally, we will study conditions under which our method blends into an undamped method, close to x_*. Let δx be some directional minimizer of the model function m_x^ω. If H_x is elliptic, then a minimizer Δx of the quadratic problem $\min f_x' v + 1/2 H_x(v,v)$ with respect the search direction δx is well defined. It can be interpreted as an undamped step, or as the step that results from the choice $\omega = 0$.

The quotient

$$\lambda := \frac{\|\delta x\|}{\|\Delta x\|} \le 1$$

can be interpreted as a damping factor. In fact, by definition we have $\delta x = \lambda \Delta x$.

For the following we will only need a slightly weaker version of the upper bound of (6):

$$x_k \to x_* \quad \text{implies} \quad \lim_{k\to\infty} \frac{R_{x_k}(\delta x_k)}{\|\delta x_k\|^2} = 0, \tag{55}$$

which is, however, stronger than (36), our minimal requirement for global convergence.

Lemma 4.10. *Let x_k be any sequence of iterates, such that H_{x_k} are uniformly elliptic. Then*

$$\lim_{k\to\infty} \frac{\omega_k R_{x_k}(\delta x_k)}{\|\delta x_k\|^2} = 0 \quad \Rightarrow \quad \lim_{k\to\infty} \lambda_k = 1.$$

Proof. To show the above equivalence we insert δx_k and Δx_k into (11) and set

$$\gamma_k := \frac{H_x(\delta x_k, \delta x_k)}{\|\delta x_k\|^2} = \frac{H_{x_k}(\Delta x_k, \Delta x_k)}{\|\Delta x_k\|^2}.$$

We obtain from (11) (with $\omega = 0$ for Δx):

$$\begin{aligned} \|\delta x_k\| \left(\frac{\omega_k}{2} \frac{R_{x_k}(\delta x_k)}{\|\delta x_k\|^2} + \gamma_k \right) &\overset{(11)}{=} |f_{x_k}' \delta x_k| / \|\delta x_k\| \\ &= |f_{x_k}' \Delta x_k| / \|\Delta x_k\| \underset{\omega=0}{\overset{(11)}{=}} \|\Delta x_k\| \gamma_k \end{aligned}$$

By assumption, the sequence γ_k is positive and bounded away from 0 and thus we obtain by division

$$1 \ge \lambda_k = \frac{\|\delta x_k\|}{\|\Delta x_k\|} = \frac{\gamma_k}{\frac{\omega_k}{2} \frac{R_{x_k}(\delta x_k)}{\|\delta x_k\|^2} + \gamma_k}$$

The right hand side tends to 1, if $\frac{\omega_k R_{x_k}(\delta x_k)}{\|\delta x_k\|^2} \to 0$. □

The following result is an immediate consequence:

Corollary 4.11. *Let x_k be a converging sequence of iterates, such that H_{x_k} are uniformly elliptic, and suppose that* (55) *holds. If ω_k is bounded, then* $\lim_{k\to\infty} \lambda_k = 1$.

To show boundedness of ω_k we consider the acceptance indicators η_k as defined in (22) and show that they tend to 1 asymptotically if the quadratic model is really a second order approximation of f in the sense of (9):

$$\lim_{k\to\infty} \frac{w_{x_k}(\delta x_k)}{\|\delta x_k\|^2} = 0.$$

It can be shown that such a condition holds, if f is twice continuously differentiable in a neighborhood of x_* and $H_x = f_x''$.

Proposition 4.12. *Suppose that $x_k \to x_*$ and assume that the second order approximation error estimate* (9) *holds. Then, independently of the choice of $\omega_k \geq 0$ we conclude for η_k, defined in* (22)*:*

$$\liminf_{k\to\infty} \eta_k \geq 1$$

for any corresponding sequence of directional minimizers δx_k.

Proof. Since, by assumption $x_k \to x_*$, we also have $\|\delta x_k\| \to 0$. Thus, by (9) we conclude

$$\lim_{k\to\infty} \frac{w_{x_k}(\delta x_k)}{\|\delta x_k\|^2} = 0, \quad \text{while by (53) we have} \quad \frac{m_{x_k}^{\omega_k}(\delta x_k)}{\|\delta x_k\|^2} \leq -\frac{\gamma}{2}.$$

Thus, taken together, we obtain

$$\lim_{k\to\infty} \frac{w_{x_k}(\delta x_k)}{m_{x_k}^{\omega_k}(\delta x_k)} = 0.$$

Hence, by definition (recall that $m_{x_k}^{\omega_k}(\delta x_k) < 0$)

$$\begin{aligned}\liminf_{k\to\infty} \eta_k = \liminf_{k\to\infty} \frac{f(x_k+\delta x_k) - f(x_k)}{m_{x_k}^{\omega_k}(\delta x_k)} &\geq \liminf_{k\to\infty} \frac{m_{x_k}^{\omega_k}(\delta x_k) - \frac{\omega_k}{6} R_{x_k}(\delta x_k) + w_{x_k}(\delta x_k)}{m_{x_k}^{\omega_k}(\delta x_k)} \\ &\geq \lim_{k\to\infty} \left(1 + \frac{w_{x_k}(\delta x_k)}{m_{x_k}^{\omega_k}(\delta x_k)}\right) = 1.\end{aligned}$$

□

Taking all facts together, we obtain our final result:

Theorem 4.13. *In addition to Assumption 4.7 suppose that* (55) *and* (9) *hold in U along x_k generated by our algorithm. If x_k comes sufficiently close to x_* then $x_k \to x_*$ and $\lambda_k \to 1$.*

Proof. By Theorem 4.9 we conclude that $x_k \to x_*$ and $\|\delta x_k\| \to 0$. By Proposition 4.12 eventually every trial correction is accepted with some $\eta_k > \overline{\eta}$. Hence, by our algorithmic restriction (27) ω_k is not increased anymore and it follows that ω_k is bounded above. This and (55), inserted into Lemma 4.10 yield the desired result. □

References

[1] Michele Benzi. Preconditioning techniques for large linear systems: a survey. *J. Comput. Phys.*, 182(2):418–477, 2002.

[2] Coralia Cartis, Nicholas Gould, and Philippe Toint. Adaptive cubic regularisation methods for unconstrained optimization. part i: motivation, convergence and numerical results. *Mathematical Programming*, pages 1–51, 2009. 10.1007/s10107-009-0286-5.

[3] Coralia Cartis, Nicholas Gould, and Philippe Toint. Adaptive cubic regularisation methods for unconstrained optimization. part ii: worst-case function- and derivative-evaluation complexity. *Mathematical Programming*, pages 1–25, 2010. 10.1007/s10107-009-0337-y.

[4] A.R. Conn, N.I.M. Gould, and P.L. Toint. *Trust-Region Methods.* SIAM, 2000.

[5] P. Deuflhard and M. Weiser. Local inexact Newton multilevel FEM for nonlinear elliptic problems. In M.-O. Bristeau, G. Etgen, W. Fitzgibbon, J.-L. Lions, J. Periaux, and M. Wheeler, editors, *Computational science for the 21st century*, pages 129–138. Wiley, 1997.

[6] P. Deuflhard and M. Weiser. Global inexact Newton multilevel FEM for nonlinear elliptic problems. In W. Hackbusch and G. Wittum, editors, *Multigrid Methods V*, Lecture Notes in Computational Science and Engineering, pages 71–89. Springer, 1998.

[7] Peter Deuflhard and Martin Weiser. *Adaptive numerical solution of PDEs*. de Gruyter Textbook. Walter de Gruyter & Co., Berlin, 2012.

[8] D. Gilbarg and N.S. Trudinger. *Elliptic Partial Differential Equations of Second Order*. Springer, 1977.

[9] A. Griewank. The modification of Newton's method for unconstrained optimization by bounding cubic terms. Technical Report NA/12, Department of Applied Mathematics and Theoretical Physics, University of Cambridge, 1981.

[10] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich. *Optimization with PDE constraints*, volume 23 of *Mathematical Modelling: Theory and Applications*. Springer, New York, 2009.

[11] J. L. Lions. *Optimal control of systems governed by partial differential equations*, volume 170 of *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag, 1971.

[12] E. Schechter. *Handbook of Analysis and its Foundations*. Academic Press, 1997.

[13] F. Tröltzsch. *Optimale Steuerung partieller Differentialgleichungen. Theorie, Verfahren und Anwendungen*. Vieweg, 2005.

[14] Fredi Tröltzsch. *Optimal control of partial differential equations*, volume 112 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2010. Theory, methods and applications, Translated from the 2005 German original by Jürgen Sprekels.

[15] M. Ulbrich. *Semismooth Newton Methods for Variational Inequalities and Constrained Optimization Problems in Function Spaces*. MOS-SIAM Series on Optimization 11. SIAM, 2011.

[16] M. Weiser, P. Deuflhard, and B. Erdmann. Affine conjugate adaptive newton methods for nonlinear elastomechanics. *Opt. Meth. Softw.*, 22(3):413–431, 2007.

[17] J. Carsten Ziems and Stefan Ulbrich. Adaptive multilevel inexact SQP methods for PDE-constrained optimization. *SIAM J. Optim.*, 21(1):1–40, 2011.