CONVERGENCE OF THE GRADIENT METHOD FOR
ILL-POSED PROBLEMS

Stefan Kindermann*
Industrial Mathematics Institute
Johannes Kepler University Linz
Altenbergerstrasse 69, 4040 Linz, Austria

(Communicated by Otmar Scherzer)

Abstract. We study the convergence of the gradient descent method for solving ill-posed problems where the solution is characterized as a global minimum of a differentiable functional in a Hilbert space. The classical least-squares functional for nonlinear operator equations is a special instance of this framework, and the gradient method then reduces to Landweber iteration. The main result of this article is a proof of weak and strong convergence under new nonlinearity conditions that generalize the classical tangential cone conditions.

1. Introduction. A widely-used approach for dealing with a nonlinear ill-posed problem is to phrase it as an operator equation in Banach- or Hilbert spaces and apply an iterative regularization method for its solution [10]. The simplest, though not the fastest, amongst them is the Landweber iteration, which can be viewed as a gradient descent method for the associated least-squares functional. A well-known convergence theory has been established for Landweber iteration for nonlinear ill-posed problems based on the seminal paper by Hanke, Neubauer, and Scherzer [13]. The pivotal innovation that paved the way for the analysis is to include appropriate restrictions on the “nonlinearity” of the problem by imposing so-called nonlinearity conditions on the underlying operators.

Such conditions have been verified for several important problems, e.g., for parameter identification in partial differential equations using interior measurements; see, e.g., [19]. However, and this is the crucial point, they have not yet been verified for certain well-studied problems like electrical impedance tomography (aka. Calderón’s problem) [9]—although Landweber and other iterative methods have been successfully applied to them. This might give a hint that the traditionally used nonlinearity conditions are too strong to be satisfied for certain applications, and one may try to replace them by weaker assumptions.

The main goal of this paper is to prove (local) weak and strong convergence of a subsequence of the gradient descent iterates for a functional with Lipschitz-continuous gradient imposing more general nonlinearity conditions than the usual ones.

Reviewing such typically-used restrictions reveals that the most common versions are the weak and strong form of the tangential cone conditions [13] [24] [26]. Stronger than those are the range-invariant conditions [13] [20]. Conceptually similar to

2010 Mathematics Subject Classification. Primary: 65J20, 65J15; Secondary: 47J06.
Key words and phrases. Inverse problems, iterative Regularization, Landweber iteration, tangential cone condition, quasiconvexity.

703 ©2017 American Institute of Mathematical Sciences
this is the approach via Hilbert scales [20, 21]. Another typical (functional-based) nonlinearity condition is the assumption of convexity, or equivalently, monotonicity of the gradient [7, 8, 23, 13, 16]. (Sometimes even strong monotonicity [22, 5] is imposed). An insightful comparison of tangential cone conditions and several versions of monotonicity of the gradient can be found in [24]; cf. also Sections 2.1 and 2.2.

We note that such conditions are also used for proving convergence of other iterative regularization methods in the nonlinear case such as, e.g., Gauss-Newton-type iterations [18, 5, 6, 17], the Levenberg-Marquardt scheme [12], or Kaczmarz iterations [11].

One of the main contribution in this article is to prove boundedness and weak convergence of the gradient descent iterations essentially under a two-parametric nonlinearity condition, which generalizes and includes both the weak tangential conditions and several convexity conditions as special cases. This is interesting insofar as the tangential cone conditions do not imply convexity of the associated least-squares functional, thus our analysis can be viewed as an attempt for a unification of the established nonlinearity restrictions.

We also prove strong convergence of the iterates under a novel restriction which requires the functional to be “balanced” around critical points. This can be seen as a generalization of the strong tangential cone condition. All these results hold both for the case of exact and noisy data, where in the latter, we employ a simple a-priori parameter choice.

Our setup is phrased as the problem of finding minima of general ill-posed functional rather than in the form of nonlinear operator equations, but since, of course, one can apply the least-squares idea, the classical Landweber iteration is a special instance of the gradient iteration studied here. Note that the Lipschitz-continuity of the gradient plays an essential role in our work, hence, certain Banach space variants of Landweber iterations (see, e.g., [14, 25]) in non-smooth spaces are not within the scope of this work.

Our paper is organized as follows: in Section 2 we define the gradient iteration, present the standard assumptions that we use and the novel nonlinearity conditions that we impose. We study them in detail by relating them to the traditionally used ones. In Section 3 we prove boundedness and weak convergence of the iterates to a stationary point for our setup both in the case of exact and noisy data. In Section 4, strong convergence of the iteration is proven in a similar framework.

2. Setup and nonlinearity conditions. We consider the problem of finding a solution of an ill-posed problem that is characterized as a global minimum of a certain functional. Throughout this paper, we denote by $B_{\rho}(x^*)$ a ball with center $x^*$ and radius $\rho$ in a Hilbert space. We assume given a Fréchet-differentiable, nonnegative functional

$$J : B_{\rho}(x^*) \subset X \to \mathbb{R}^+,$$

where $X$ is a Hilbert space, and that a sought-for solution $x^*$ satisfies

$$J(x^*) = 0.$$

By the nonnegativity of $J$, $x^*$ is a global minimum in $B_{\rho}(x^*)$ and has to satisfy the first-order optimality condition

$$\nabla J(x^*) = 0.$$
The most important instance of such a functional is the least-squares functional \( J_{LS}(x) \) for a nonlinear operator equation with given data \( y \),

\[
F(x) = y,
\]

which is defined as

\[
J_{LS}(x) = \frac{1}{2} \| F(x) - y \|^2.
\]

In the setup of (1), we assumed that the given data are encoded somehow into the functional \( J \). Similar to the least-squares case, we have to allow for inexact data as well, i.e., we have to consider a “noisy” version of \( J \) that represents the actual measurements.

Thus, we assume given a Fréchet-differentiable, nonnegative functional

\[
J^\delta : B_\rho(x^*) \subset X \to \mathbb{R}^+,
\]

where the actual iteration is based upon. In order to solve (2) for \( x^* \) with given noisy data, a gradient iteration can be used. It is defined iteratively (as long as the iterates \( x_k^\delta \) stay in \( B_\rho(x^*) \)) as

\[
x_{k+1}^\delta = x_k^\delta - \nabla J^\delta(x_k^\delta), \quad k = 0, \ldots,
\]

starting with an initial guess \( x_0^\delta \in B_\rho(x^*) \). For the analysis, it is convenient to define the corresponding iteration with exact data as well,

\[
x_{k+1} = x_k - \nabla J(x_k), \quad k = 0, \ldots,
\]

starting with the same initial guess \( x_0 = x_0^\delta \in B_\rho(x^*) \) as in (7). In the least-squares case (5), iteration (7), respectively (8), is the classical Landweber iteration:

\[
x_{k+1}^\delta = x_k^\delta - F^\delta(x_k^\delta) (F(x_k^\delta) - y^\delta), \quad k = 0, \ldots
\]

Note that the gradient descent iterations usually allow for a stepsize parameter in front of the gradient term. We assume a constant stepsize parameter that is encoded into the functional \( J \) such that it will be set to 1 throughout. The only restriction on the stepsize comes from the assumptions that we impose on \( J \) and \( J^\delta \).

Essentially, we assume that \( J \) and \( J^\delta \) are differentiable on \( B_\rho(x^*) \) with Lipschitz-continuous derivative and Lipschitz constant smaller than 1. Precisely, we postulate the following:

**Assumption 1.**

1. \( X \) is a Hilbert space with inner product denoted by \( \langle \cdot, \cdot \rangle \).
2. For some \( \rho > 0 \), the functionals \( J \) and \( J^\delta \) are defined on \( B_\rho(x^*) \subset X \) and are Fréchet-differentiable there.
3. There exists an \( x^* \) (exact solution) which satisfies (2) (and hence also (3)).
4. For all \( x \in B_\rho(x^*) \), the gradient \( \nabla J(x) \) is Lipschitz continuous with Lipschitz constant \( L < 1 \):

\[
\| \nabla J(x_1 + x_2) - \nabla J(x_1) \| \leq L \| x_2 \| < \| x_2 \|, \quad \forall x_1, x_1 + x_2 \in B_\rho(x^*).
\]

5. The functional \( J^\delta \) satisfies

\[
\| \nabla J^\delta(x) - \nabla J(x) \| \leq \delta \quad \forall x \in B_\rho(x^*), \quad \text{and}
\]

\[
J^\delta(x) - J(x) \leq \psi(\delta) \quad \forall x \in B_\rho(x^*), \quad \text{with} \ \lim_{\delta \to 0} \psi(\delta) = 0.
\]
6. The gradient satisfies

\[ \| \nabla J(\delta)(x) \|_2^2 \leq \phi(J(\delta)(x)) \quad \forall x \in B_\rho(x^*) \]

with some monotone positive continuous function \( \phi \) with \( \phi(0) = 0 \).

As the notation suggests, \( \delta \) plays the role of the noise level. We note that (11) implies that \( \nabla J(\delta) \) is Lipschitz continuous with Lipschitz constant \( L_\delta \)

\[ L_\delta \leq L + \delta. \]

The assumptions imposed here are all quite standard: 1.–4. in Assumptions 1 are traditionally employed for iterative regularization schemes in Hilbert spaces, cf., e.g., [13, 10, 4, 19]. A slight difference to the common setup in literature is the fact that we additionally ask for the Lipschitz constant \( L \) to be less than 1 in 4., whereas typically only \( L < \infty \) is postulated.

We would like to point out that in the context of regularization of nonlinear ill-posed problem, gradient methods are almost exclusively studied for the case of \( J \) being the least-squares functional for an operator equation. In this case, we have that \( \nabla J(\delta)(x) = F'(x^*)(F(x) - y^\delta) \), and (11), (12)–(13) follow easily from the definition of the noise-level there, \( \delta_{\text{LS}} = \| y - y^\delta \| \), and the standard condition that the Fréchet-derivative is bounded on \( B_\rho(x^*) \):

\[ \sup_{x \in B_\rho(x^*)} \| F'(x) \| \leq C \]

(cf. [13, 19, 4]). In fact, in the least-squares case, together with (15), we have that \( \delta = C\delta_{\text{LS}} \), while \( \psi(\delta) = \delta_{\text{LS}} = \frac{\delta}{C} \) and the inequality (13) holds with \( \phi(s) = Cs \). Since we study here the slightly more general setup of functional minimization, we need the new conditions (12) and (13) as a tool to link gradient and functional. All in all, we consider Assumptions 1 comparable to those in the literature.

2.1. Nonlinearity conditions. In this section, we introduce new nonlinearity conditions. We group them loosely into those sufficient for weak convergence and those for strong convergence.

2.1.1. Weak nonlinearity conditions. We propose a two-parametric generalization of the well-known weak tangential cone condition:

**Definition 2.1.** For some \( \gamma \in [0, \infty) \) and \( \beta \in \mathbb{R} \), we say that NC(\( \gamma, \beta \)) is satisfied for \( J \) if for all \( x_1, x_2 \in B_\rho(x^*) \) the following implication holds true:

\[ J(x_1) \leq \gamma J(x_2) \Rightarrow \langle \nabla J(x_2), x_2 - x_1 \rangle \geq -\beta \| \nabla J(x_2) \|^2. \]

Note that we allow \( \gamma = 0 \) and \( \gamma = \infty \). In the later case, the premise in the implication is tautological, thus the conclusion has to hold for all \( x_1, x_2 \in B_\rho(x^*) \), while for \( \gamma = 0 \), the conclusion has to hold only for \( x_1 \) at a global minimum.

It is easy to verify that the condition in Definition 2.1 is the stronger the larger the \( \gamma \) and the smaller the \( \beta \) is:

- for \( \gamma_1 \leq \gamma_2 \) : NC(\( \gamma_2, \beta \)) \Rightarrow NC(\( \gamma_1, \beta \))
- for \( \beta_1 \leq \beta_2 \) : NC(\( \gamma, \beta_1 \)) \Rightarrow NC(\( \gamma, \beta_2 \)).

Let us compare this condition with some established ones used in the convergence analysis of the gradient iteration for ill-posed problems. Quite often, in optimization, the convexity of the functional is imposed. In our setup, this is equivalent to
the monotonicity of the gradient, i.e.,
\[ \langle \nabla J(x_2) - \nabla J(x_1), x_2 - x_1 \rangle \geq 0 \quad \forall x_1, x_2 \in B_p(x^*) . \]
This condition guarantees weak convergence of the gradient method, which follows, e.g., from Scherzer’s results [24]. Monotonicity has been used in the analysis for the continuous version of Landweber iteration, see, e.g. [23, 15, 16] and also in [3] Chpt. 6 for Landweber iteration with an additional projection involved. According to our best knowledge, except for strong monotonicity, a monotone gradient (resp. a convex functional) does not suffice to prove strong convergence for the classical Landweber iteration. (Note that in [23, 15, 16] a continuous version was investigated while in [5, 7, 8] a modified (i.e., regularized) Landweber iteration was considered.)

A weaker condition than convexity is that of \( \{x^*\}\)-quasi-monotonicity, i.e.,
\[ \langle \nabla J(x_2) - \nabla J(x^*), x_2 - x^* \rangle \geq 0 \quad \forall x_2 \in B_p(x^*), \]
and this is implied by the \( \{x^*\}\)-quasi-uniform-monotonicity for some \( \alpha > 0 \)
\[ \langle \nabla J(x_2) - \nabla J(x^*), x_2 - x^* \rangle \geq \alpha \| \nabla J(x_2) - \nabla J(x^*) \| \quad \forall x_2 \in B_p(x^*). \]
These conditions can be found in [24] (compare also [8]). Since \( x^* \) is a critical point by definition, we have that \( \{17\} \) and \( \{18\} \) is equivalent to \( \text{NC}(0, 0) \) and \( \text{NC}(0, -\alpha) \), respectively. Scherzer [24, Thm. 2.12] and also Vasin [26] proved weak convergence for Landweber iteration under \( \{18\} \) with \( \alpha > 0 \) (thus for \( \text{NC}(0, \beta) \) for \( \beta < 0 \)). We will extend their results to the case \( \beta \in \mathbb{R} \).

We have seen that \( \text{NC}(0, \beta) \) can be related to various established quasi-monotonicity conditions. Next, we consider the case that \( \beta = 0 \) and \( \gamma \in [0, 1] \). In this case, the parametric nonlinearity conditions \( \text{NC}(\gamma, 0) \) can be related to the notion of quasi-convexity. At first we introduce a generalization of quasi-convexity:

**Definition 2.2.** Let \( C \) be a convex set and let \( 0 \leq \gamma \leq 1 \). We say that the functional \( f \) is \( \gamma \)-quasiconvex if for all \( x, y \in C \) and \( \lambda \in (0, 1) \) we have
\[ f(x_1) \leq \gamma f(x_2) \Rightarrow f(\lambda x_1 + (1 - \lambda)x_2) \leq f(x_2). \]

For \( \gamma = 1 \), we encounter the traditional definition of quasiconvexity [2], which might also be phrased as the condition that
\[ f(\lambda x_1 + (1 - \lambda)x_2) \leq \max\{f(x_2), f(x_1)\} \quad \forall x_1, x_2 \in C, \lambda \in (0, 1). \]
For positive functionals \( f \), the assumption of \( \gamma \)-quasiconvexity is weaker than quasiconvexity, which itself is in any case weaker than convexity. In terms of level-sets, it is easy to see that \( f \) is quasiconvex if and only if all its lower level sets \( \{f \leq \alpha\} \) are convex. We may view \( \gamma \)-quasiconvexity as the condition that the convex hull of \( \{f \leq \gamma \alpha\} \) does not intersect the complement of \( \{f \leq \alpha\} \).

The following characterization of \( \gamma \)-quasiconvexity is useful:

**Proposition 1.** Let \( f : X \to \mathbb{R} \) be Fréchet-differentiable on the open convex set \( C \).
Let \( 0 \leq \gamma \leq 1 \). Then \( f \) is \( \gamma \)-quasiconvex on \( C \) if and only if \( \text{NC}(\gamma, 0) \) holds.

**Proof.** We follow [1], cf. [2] Theorem 3.11. If \( f \) is \( \gamma \)-quasiconvex, then for any \( x_1, x_2 \) with \( f(x_1) \leq \gamma f(x_2) \), we have \( f(\lambda x_1 + (1 - \lambda)x_2) \leq \gamma (1 - \lambda) f(x_2) \leq 0 \). The limit \( \lambda \to 0 \) implies that \( \langle \nabla f(x_2), x_1 - x_2 \rangle \leq 0 \), hence \( \text{NC}(\gamma, 0) \) holds. Conversely, suppose that \( \text{NC}(\gamma, 0) \) holds, and define for \( x_1, x_2 \) with \( f(x_1) \leq \gamma f(x_2) \) the following function \( G(\lambda) := f(\lambda x_1 + (1 - \lambda)x_2) \) on the unit interval \( [0, 1] \). Suppose that \( G(\lambda) > G(0) \) for some \( \lambda \in [0, 1] \). Consider the largest element \( \xi^* \) in the nonempty closed set \( \{\xi \in [0, \lambda] | G(\xi) \leq G(0)\} \). By construction \( \xi^* > \lambda \), and it follows by the intermediate
value theorem that there exists a $\xi \in (\zeta^*, \lambda)$ with $G''(\xi) > 0$ and $G(\xi) > G(0)$ (since $\zeta^*$ is the largest element in the set). However, then with $x_\xi = \xi x_1 + (1 - \xi) x_2$ we find that $f(x_1) = G(1) \leq \gamma f(x_2) = \gamma G(0) < \gamma f(x_\xi)$. Thus, NC($\gamma, 0$) implies that

$$0 \leq \langle \nabla f(x_\xi), x_\xi - x_1 \rangle = (1 - \xi) \langle \nabla f(x_\xi), x_2 - x_1 \rangle = -(1 - \xi) G'(\xi),$$

which contradicts $G'(\xi) > 0$. Hence, $G(\lambda) \leq G(0)$ must hold, which implies $\gamma$-quasiconvexity.

In particular, quasiconvexity is equivalent to NC(1, 0).

### 2.1.2. Strong nonlinearity conditions.

Strong convergence of gradient iterations require a stronger nonlinearity condition than the previous ones. For our convergence analysis, we need additionally to (16) the following one, which we denote as balancing condition.

**Definition 2.3.** Let $\gamma \geq 0$. We say that the functional $J : B_\rho(x^*) \to \mathbb{R}^+$ is $\gamma$-balanced around $x^*$ if for some $\rho_0 < \rho$ and any sequence $z_n$ with $\rho_0 \leq \|z_n\| \leq \rho$, there exists a $\tau > 0$ and a $n_0 \in \mathbb{N}$ such that

$$J(x^* - \tau z_n) \leq \gamma J(x^* + z_n) \quad \forall n \geq n_0. \tag{21}$$

We will prove strong convergence of the gradient iteration under the condition that $J$ is $\gamma$-balanced and satisfies NC($\gamma, \beta$) for some $\gamma \geq 0$ and some $\beta \in \mathbb{R}$.

The condition in Definition 2.3 can sloppily be interpreted as the requirement that $J$ does not have extremely large values when evaluated at a mirror point around $x^*$. Thus, the functional should roughly behave in a similar way left and right at $x^*$ on a line through $x^*$.

It is easy to verify that if $J$ is convex on $B_\rho(x^*)$ and satisfies a symmetry condition

$$J(x^* - z) \leq C J(x^* + z) \quad \forall z \in B_\rho(x^*),$$

with a constant $C$, then (21) holds.

### 2.2. The least-squares case.

In case that $J$ is defined as a least squares functional $F$ based on a nonlinear operator $F$, the functional-based nonlinearity condition above can be expressed as nonlinearity conditions for the operator.

The most popular nonlinearity condition ensuring weak convergence for $J_{LS}$ is the weak tangential cone condition [24]: there exists an $0 < \eta < 1$ such that

$$\langle F(x) - F(x^*) - F'(x)(x - x^*), F(x) - F(x^*) \rangle \leq \eta \| F(x) - F(x^*) \|^2 \quad \forall x \in B_\rho(x^*), \tag{22}$$

or, equivalently,

$$\|F'(x)(x - x^*), F(x) - F(x^*)\| \geq (1 - \eta) \| F(x) - F(x^*) \|^2 \quad \forall x \in B_\rho(x^*), \tag{23}$$

or, equivalently,

$$\langle \nabla J_{LS}(x_2), x_2 - x^* \rangle \geq (1 - \eta) J_{LS}(x_2) \quad \forall x_2 \in B_\rho(x^*). \tag{24}$$

Thus, the weak tangential cone condition is almost (up to replacing $\|\nabla J_{LS}(x_2)\|^2$ by $J(x_2)$) the $\{x^*\}$-quasi-uniform-monotonicity (cf. [24] Thm 3.4)). In fact, for Fréchet-differentiable $F$, (22) with $\eta \in (0, 1)$ implies NC(0, $\beta$) with a negative $\beta < 0$. It was shown [24] that (22) with $\eta \in (0, 1)$ implies weak convergence for the Landweber iteration with exact data. We generalize these results insofar as we also verify weak convergence in the noisy case and more interesting, we prove that (16) with $\gamma = 0$ and any $\beta \in \mathbb{R}$ (also positive ones and in particular for (22) with $\eta = 1$) already implies weak convergence of a subsequence of the gradient iteration.
Traditionally, strong convergence of the Landweber iteration is verified under the so-called (strong) tangential cone condition (or strong Scherzer condition) (see, e.g., [13, 10, 19, 24]): there exists $0 < \eta < 1$ such that

$$\|F(x) - F(\hat{x}) - F'(x)(x - \hat{x})\| \leq \eta \|F(x) - F(\hat{x})\| \quad \forall x, \hat{x} \in B_\rho(x^*),$$

(25)
or equivalently, by squaring and expanding the square on the left,

$$\langle F'(x)(x - \hat{x}), F(x) - F(\hat{x}) \rangle$$

$$\geq (1 - \eta^2)\frac{1}{2} \|F(x) - F(\hat{x})\|^2 + \frac{1}{2}\|F'(x)(x - \hat{x})\|^2 \quad \forall x, \hat{x} \in B_\rho(x^*).$$

(26)

It is obvious that the strong tangential cone condition implies the weak one. There are several interesting conclusions that follow from (25), for instance [19], it postulates the existence of a family of operators $R_\rho$ such that

$$F'(x) = R_\rho F'(x^*) \quad \forall x \in B_\rho(x^*) \quad \text{and} \quad \|R_\rho - I\| \leq C \|x - x^*\|.$$

Locally, it follows that $R_\rho$ are invertible operators, which allows to compare the derivatives at different points $x$ such that this condition implies (25) (possibly on a smaller ball).

It is interesting that the weak tangential cone condition can also be expressed by a derivative-free condition:

**Proposition 2.** Let $F$ be Fréchet-differentiable. Then, condition (22) holds for some $\eta \in [0,1]$ if and only if the least-squares functional $J_{LS}$ has the property that the mapping

$$t \to \frac{1}{t^{2(1-\eta)}} J_{LS}(x^* + t(x - x^*))$$

(28)
is monotonically increasing for $t \in [0,1]$ for all $x \in B_\rho(x^*)$.

**Proof.** Let $x_t := x^* + t(x - x^*)$ and let $G(t) := J_{LS}(x^* + t(x - x^*))$. Since

$$G'(t) = \langle F'(x_t)(x - x^*), F(x_t) - F(x^*) \rangle = \frac{1}{2}\langle F'(x_t)(x_t - x^*), F(x_t) - F(x^*) \rangle,$$

we have that (22) implies that $G'(t) \geq (1 - \eta)2\sqrt{2} G(t)$, from which the monotonicity of (28) follows by calculating the derivative. On the other hand, if (28) is monotone, then $G'(t) \geq (1 - \eta)2\sqrt{2} G(t)$ for $t \in (0,1)$ follows easily, and hence, taking $t = 1$, we obtain (22).

Since we motivated our analysis by the claim that the new nonlinearity conditions are more general than traditional ones, we next verify that the strong tangential cone condition implies NC($\gamma, \beta$) with appropriate parameter values and also the balancing condition.

**Lemma 2.4.** Let the tangential cone condition (25) hold with $\eta < 1$. Then the least-squares functional $F$ satisfies NC($\gamma, \beta$) with parameter $\gamma < \left(\frac{\sqrt{1-\eta^2}}{1+\sqrt{1-\eta^2}}\right)^2 < 1$ and $\beta = -[(1 - \eta^2)(1 - \sqrt{\gamma})^2 - \gamma] \frac{1}{2(\sup_{x \in B_\rho(x^*)}\|F'(x^*)\|)^2} < 0$. 

**Inverse Problems and Imaging** Volume 11, No. 4 (2017), 703–720
Lemma 2.6. Let the classical weak tangential cone condition holds.

Proof. Assume that \( J(x_1) \leq \gamma J(x_2) \). Using \( \alpha(x) \) with \( x = x_2, \hat{x} = x_1 \), Young's inequality, and the triangle inequality, we obtain

\[
\langle \nabla J_{LS}(x_2), (x_2 - x_1) \rangle = \langle F(x_2) - F(x^*), F'(x_2)(x_2 - x_1) \rangle
\]

\[
= (F(x_2) - F(x_1), F'(x_2)(x_2 - x_1)) + (F(x_1) - F(x^*), F'(x_2)(x_2 - x_1))
\]

\[
\geq (1 - \eta^2) \frac{1}{2} \|F(x_2) - F(x_1)\|^2 + \frac{1}{2} \|F'(x_2)(x_2 - x_1)\|^2
\]

\[
- \|F'(x_2)(x_2 - x_1)\| \|F(x_1) - F(x^*)\|
\]

\[
\geq (1 - \eta^2) \frac{1}{2} \|F(x_2) - F(x_1)\|^2 - \frac{1}{2} \|F(x_1) - F(x^*)\|^2
\]

\[
\geq (1 - \eta^2)(1 - \sqrt{\gamma}) \frac{1}{2} \|F(x_2) - F(x^*)\|^2 - \frac{1}{2} \|F(x_2) - F(x^*)\|^2
\]

\[
\geq [(1 - \eta^2)(1 - \sqrt{\gamma}) - \gamma] J(x_2) \geq [(1 - \eta^2)(1 - \sqrt{\gamma}) - \gamma] C^{-2} \frac{1}{2} \|\nabla J(x_2)\|^2,
\]

with \( C = \sup_{x \in B_p(x^*)} \|F'(x^*)\| \).

This lemma justifies our claim that NC(\( \gamma, \beta \)) is a generalization of the tangential cone condition. Note, however, that quasiconvexity (i.e., NC(1, 0)) or even convexity of the least-squares functional is not implied by the tangential cone conditions, while by our results, strong convergence holds for quasiconvex (and convex) functionals if the balancing condition is additionally satisfied.

Concerning the balancing condition, it can be shown that the least-squares functional is \( \gamma \)-balanced if \( F \) satisfies the strong tangential cone condition.

Lemma 2.5. Let (25) hold with \( \eta < 1 \). Then the least-squares functional (5) is \( \gamma \)-balanced around \( x^* \) for any \( \gamma \in (0, 1) \).

Proof. Indeed it follows from (27) that

\[
2J(x^* - \tau z) = \|F(x^* - \tau z) - F(x^*)\|^2 \leq \frac{2}{1 - \eta} \|F'(x^*)(-\tau z)\|^2
\]

\[
= \tau \frac{2}{1 - \eta} \|F'(x^*)z\|^2 \leq \tau^2 \frac{2(1 + \eta)}{1 - \eta} \|F(x^* + z) - F(x^*)\|^2
\]

\[
= \tau \frac{1 + \eta}{(1 - \eta)\gamma} 2J(x^* + z).
\]

Thus \( \tau = \frac{1 - \eta}{(1 + \eta)\gamma} \) provides (21).

We note that for convergence of the Landweber iteration, often the tangential cone condition is imposed with \( \eta < \frac{1}{2} \). A consequence of our results is that strong convergence also follows with \( \eta < 1 \).

We provide another sufficient condition for the balancing condition (21) if the classical weak tangential cone condition holds.

Lemma 2.6. Let \( x^* \) be the unique global minimum in \( B_p(x^*) \) and let the weak tangential cone condition (22) hold for some \( \eta < 1 \). If for any sequence with \( \Delta_n \to 0 \)

\[
\limsup_n \frac{J_{LS}(x^* + \Delta_n)}{J_{LS}(x^* - \Delta_n)} > 0,
\]

holds, then (21) is satisfied for any \( \gamma > 0 \).
Proof. Fix $\gamma > 0$ and suppose that (21) does not hold. Then we find a sequence $z_n$ with $\rho_0 \leq \|z_n\| \leq \rho$ and a sequence of $\tau_n > 0$ with $\tau_n \to \infty$ and

$$J_{LS}(x^* - \tau_n z_n) > \gamma J(x^* + z_n).$$

However, Proposition 2 implies that $J_{LS}(x^* + z_n) \geq \tau_n^{-2(1-\eta)} J_{LS}(x^* + \tau_n z_n)$. Since $\Delta_n := \tau_n z_n \to 0$, we obtain

$$J_{LS}(x^* + \Delta_n) < \frac{\tau_n^{-2(1-\eta)}}{\gamma} \to 0,$$

which contradicts (29).

This also illustrates that for linear problems, the balancing conditions is trivial. Indeed, as the functional $J(x^* + \Delta)$ is a quadratic form $(A\Delta, \Delta)$, then, the ratio in (29) is always 1. If $J$ can be estimated around $x^*$ from below and above by a constant times an even-homogeneous functional (similar to (27)), then (29) is satisfied.

As a justification for our claim of a unification of nonlinearity conditions, we present the implications of these conditions in the following scheme. The main result about weak convergence in this paper (Section 3) is indicated in the last line of this table.

| convexity | $\Rightarrow$ quasiconvexity $\Leftrightarrow$ NC($\gamma$,0) |
|-----------|-----------------------------------------------------------|
| weak tangential cone cond. $\uparrow$ | $\Rightarrow$ NC(0,$\beta < 0$) $\Rightarrow$ NC(0,0) $\Rightarrow$ NC(0,$\beta > 0$) $\downarrow$ weak convergence |

3. **Weak convergence.** For the following analysis, it is convenient to introduce some shorthand notations both for the noisy and exact case:

$$e_k := x_k - x^*, \quad \delta_k := x_k - x^*,$$

$$\nabla J_k := \nabla J(x_k), \quad \nabla J_\delta := \nabla J(\delta).$$

The gradient iterations can then be written as

$$e_{k+1}^\delta = e_k^{\delta} - \nabla J_k^\delta, \quad e_{k+1} = e_k - \nabla J_k.$$

The first lemma concerns monotonicity of the functional values.

**Lemma 3.1.** Let Assumption [1] hold and let $x_k, x_{k+1} \in B_\rho(x^*)$ be defined by (8). Then the functional values are monotonically decreasing:

$$J(x_{k+1}) \leq J(x_k).$$

Moreover, if $x_k \in B_\rho(x^*)$ for $k = 0, \ldots, N$, then

$$\sum_{k=0}^{N-1} \|x_{k+1} - x_k\|^2 = \sum_{k=0}^{N-1} \|\nabla J_k\|^2 \leq \frac{1}{(1-L)} J(x_0) < \infty.$$

**Proof.** By Lipschitz continuity and (10), we have using $\Delta_k = x_{k+1} - x_k$,

$$|J(x_k + \Delta_k) - J(x_k) - (\nabla J(x_k), \Delta_k)| \leq L\|\Delta_k\|^2.$$

By (3) we have

$$\langle \nabla J(x_k), \Delta_k \rangle = -\|\nabla J_k\|^2 = -\|\Delta_k\|^2.$$
Thus with (34) we obtain
\begin{equation}
J(x_{k+1}) = J(x_k) + \Delta_k = J(x_k) - (L - 1)\|\Delta_k\|^2 < 0,
\end{equation}
which proves the first assertion. A telescope sum,
\begin{equation*}
J(x_N) - J(x_0) + (1 - L) \sum_{i=0}^{N-1} \|\nabla J_i\|^2 < 0,
\end{equation*}
yields the second result.

By completely the same proof and by replacing \( J \) by \( J^\delta \) and using the “noisy” variables instead of the exact ones, we can verify the analogous result for \( J^\delta \).

**Lemma 3.2.** Let Assumption 1 hold and let \( x_k^\delta \) be defined by (34) and assume that \( x_k^\delta, x_{k+1}^\delta \in B_p(x^*) \) and that the Lipschitz constant of \( J^\delta \) satisfies \( L^\delta < 1 \). Then the corresponding residuals are monotonically decreasing:
\begin{equation}
J^\delta(x_{k+1}^\delta) \leq J^\delta(x_k^\delta).
\end{equation}
Moreover, if \( x_k^\delta \in B_p(x^*) \) for \( k = 0, \ldots, N \), then
\begin{equation*}
\sum_{k=0}^{N-1} \|x_{k+1}^\delta - x_k^\delta\|^2 = \sum_{k=0}^{N-1} \|\nabla J_k^\delta\|^2 \leq \frac{1}{(1 - L^\delta)} J^\delta(x_0) < \infty.
\end{equation*}

Next, we consider uniform bounds for the error for the iteration with the exact functional \( J \). We recall the definition of the positive part \( f^+ := \max(f, 0) \).

**Lemma 3.3.** Let Assumption 1 hold. Suppose that \( x_k \in B_p(x^*) \), for \( k = 0, \ldots, N \). Assume that NC(0, \( \beta \)) holds for some \( \beta \in \mathbb{R} \). Then
\begin{equation*}
\|e_{k+1}\|^2 \leq \|e_0\|^2 + \frac{(1 + 2\beta)^+}{(1 - L)} J(x_0), \quad k = 0, \ldots, N.
\end{equation*}

**Proof.** By (32) and with (16), we have for \( k \leq N \)
\begin{align*}
\|e_{k+1}\|^2 &= \|e_k\|^2 - 2(\nabla J_k, e_k) + \|\nabla J_k\|^2 \\
&\leq \|e_k\|^2 + 2\beta \|\nabla J_k\|^2 + \|\nabla J_k\|^2 \\
&= \|e_k\|^2 + (1 + 2\beta) \|\nabla J_k\|^2.
\end{align*}
By telescoping we find with Lemma 3.1
\begin{equation*}
\|e_{k+1}\|^2 - \|e_0\|^2 \leq (1 + 2\beta) \sum_{l=0}^{k-1} \|\nabla J_l\|^2 \leq \frac{(1 + 2\beta)^+}{(1 - L)} J(x_0).
\end{equation*}

This lemma gives boundedness of the exact Landweber iteration.

**Corollary 1.** Let Assumption 1 and let NC(0, \( \beta \)) hold for some \( \beta \in \mathbb{R} \). Suppose that \( x_0 \) is such that
\begin{equation}
\|x_0 - x^*\|^2 + \frac{(1 + 2\beta)^+}{(1 - L)} J(x_0) < \rho^2.
\end{equation}
Then \( x_k \in B_p(x^*) \) for all \( k \geq 0 \).

**Proof.** We proceed by induction. Clearly \( x_0 \in B_p(x^*) \) by (37). Suppose that \( x_l \in B_p(x^*) \) for all \( 0 \leq l \leq k \). Then Lemma 3.3 with \( N = k \) yields that \( \|e_{k+1}\| < \rho \), thus, \( x_{k+1} \in B_p(x^*) \). By induction it follows that \( x_k \in B_p(x^*) \) for all \( k \geq 0 \). \( \square \)
Next, we consider the noisy iteration and verify a uniform bound for \( e_k^\delta \). The first lemma provides a recursive estimate.

**Lemma 3.4.** Let Assumption [\ref{assumption}](#assumption) hold. Suppose that \( x_k^\delta \in B_\rho(x^*) \) and let NC(0, \( \beta \)) hold for some \( \beta \in \mathbb{R} \). Then

\[
\| e_{k+1}^\delta \|^2 \leq \| e_k^\delta \|^2 + \| \nabla J_k^\delta \|^2 \theta + 4 \beta^+ \delta^2,
\]

with

\[
\theta = (1 + 4 \beta^+).
\]

**Proof.** Define \( H_k^\delta = \nabla J(x_k^\delta) - \nabla J(x_k^\delta) \). We obtain with the help of (16), (11), and Young’s inequality,

\[
\| e_{k+1}^\delta \|^2 = \| e_k^\delta \|^2 + \| \nabla J_k^\delta \|^2 - 2 \langle \nabla J_k^\delta, e_k^\delta \rangle
\]

\[
\leq \| e_k^\delta \|^2 + \| \nabla J_k^\delta \|^2 - 2 \langle \nabla J(x_k^\delta), e_k^\delta \rangle - 2 \langle H_k^\delta, e_k^\delta \rangle
\]

\[
\leq \| e_k^\delta \|^2 + \| \nabla J_k^\delta \|^2 + 2 \beta \| \nabla J(x_k^\delta) \|^2 + 2 \| e_k^\delta \| \| H_k^\delta \|
\]

\[
\leq \| e_k^\delta \|^2 + \| \nabla J_k^\delta \|^2 + 2 \beta \| \nabla J_k^\delta \|^2 + 2 \beta \| H_k^\delta \|^2
\]

\[
+ 4 \beta \| \nabla J_k^\delta \| \| H_k^\delta \| + 2 \| e_k^\delta \| \| H_k^\delta \|
\]

\[
\leq \| e_k^\delta \|^2 + \| \nabla J_k^\delta \|^2 (1 + 2 \beta + 2 \| \beta \|) + (2 \| \beta \| + 2 \beta) \| H_k^\delta \|^2
\]

\[
+ 2 \| H_k^\delta \| \| e_k^\delta \|.
\]

With (11), the inequality (38) follows. \( \square \)

The next lemma provides a uniform bound.

**Lemma 3.5.** Let Assumption [\ref{assumption}](#assumption) hold and let \( L_\delta < 1 \). Suppose that \( x_k^\delta \in B_\rho(x^*) \) for \( k = 0, \ldots, N \), and let NC(0, \( \beta \)) hold for some \( \beta \in \mathbb{R} \). Define

\[
\xi = \max\{1, 2 \sqrt{\beta^+}\}.
\]

Then

\[
\| e_{k+1}^\delta \|^2 + \frac{\theta}{1 - L_\delta} J(x_0) - \sum_{i=0}^k \| \nabla J_k^\delta \|^2 \theta
\]

\[
\leq \left( \sqrt{\| e_0^\delta \|^2 + \frac{\theta}{1 - L_\delta} J(x_0) + \xi \delta k} \right)^2 \quad k = 0, \ldots, N.
\]

**Proof.** We proceed by induction over \( N \). Let \( N = 0 \) and assume that \( x_0 \in B_\rho(x^*) \).

For \( k = 0 \) we have by (38)

\[
\| e_1^\delta \|^2 + \frac{\theta}{1 - L_\delta} J(x_0) - \| \nabla J_0^\delta \|^2 \theta
\]

\[
\leq \| e_0^\delta \|^2 + \| \nabla J_0^\delta \|^2 \theta + 2 \delta \| e_0^\delta \| + 4 \beta^+ \delta^2 + \frac{\theta}{1 - L_\delta} J(x_0) - \| \nabla J_0^\delta \|^2 \theta
\]

\[
\leq \| e_0^\delta \|^2 + 2 \delta \| e_0^\delta \| + 4 \beta^+ \delta^2 + \frac{\theta}{1 - L_\delta} J(x_0)
\]
Proposition 3. Let Assumption 1 and

\[ \left( \sqrt{\| e_0^\delta \|^2 + \frac{\theta}{1 - L_\delta} J^\delta(x_0)} \right)^2 + 2\xi \delta \| e_0^\delta \| + 4\beta^+ \delta^2 \]

\leq \left( \sqrt{\| e_0^\delta \|^2 + \frac{\theta}{1 - L_\delta} J^\delta(x_0) + \xi \delta} \right)^2 + \delta^2(4\beta^+ - \xi^2).

Since the last term is negative by definition of \( \xi \), the estimate holds for \( k = 0 = N \).

Now suppose that if \( x^\delta_k \in B_\rho(x^*) \) for \( k = 0, \ldots, N - 1 \), then the estimate (40) holds for \( k = 0, \ldots, N - 1 \). We show that this is also the case when \( N \) is replaced by \( N + 1 \). Thus, let \( x^\delta_k \in B_\rho(x^*) \) for \( k = 0, \ldots, N \). By the induction hypothesis we only have to show that (40) holds for \( k = N \).

By Lemma 3.2, we obtain

\begin{equation}
\sum_{l=0}^{N-1} \| \nabla J^\delta_l \|^2 \leq \frac{1}{(1 - L_\delta)} J^\delta(x_0).
\end{equation}

For brevity, define \( \kappa = \frac{\theta}{(1 - L_\delta)} J^\delta(x_0) \). By Lemma 3.4 and since \( \xi \geq 1 \), we find

\[ \| e_{N+1}^\delta \|^2 - \theta \sum_{l=0}^{N} \| \nabla J^\delta_l \|^2 + \kappa \leq \| e_N^\delta \|^2 - \theta \sum_{l=0}^{N-1} \| \nabla J^\delta_l \|^2 + \kappa + 2\delta \| e_{N-1}^\delta \| + 4\beta^+ \delta^2 \]

\[ \leq \| e_N^\delta \|^2 - \theta \sum_{l=0}^{N-1} \| \nabla J^\delta_l \|^2 + \kappa + 2\xi \delta \| e_N^\delta \| + 4\beta^+ \delta^2. \]

According to the induction hypothesis we have (40), which allows to estimate the first three terms on the right-hand side. Moreover, by (41) and (40), again \( \| e_N^\delta \| \) can be bounded by the right-hand side in (40). Thus

\[ \| e_{N+1}^\delta \|^2 - \theta \sum_{l=0}^{N} \| \nabla J^\delta_l \|^2 + \kappa \leq \left( \sqrt{\| e_0^\delta \|^2 + \frac{\theta}{(1 - L_\delta)} J^\delta(x_0) + \xi \delta(N - 1)} \right)^2 + 2\delta \xi \left( \sqrt{\| e_0^\delta \|^2 + \frac{\theta}{(1 - L_\delta)} J^\delta(x_0) + \xi \delta(N - 1)} \right) + 4\beta^+ \delta^2. \]

By completing the square as before and since \( (4\beta^+ - \xi^2) \leq 0 \) we find (40) for \( k \leq N \), which proves the lemma.

We have the following proposition:

**Proposition 3.** Let Assumption 2 and NC(0, \( \beta \)) hold for some \( \beta \in \mathbb{R} \). Let \( L_\delta < 1 \) in \( B_\rho(x^*) \) and \( x_0^\delta \) and \( N \geq 0 \) be such that

\begin{equation}
\left( \sqrt{\| e_0^\delta \|^2 + \frac{\theta}{(1 - L_\delta)} J^\delta(x_0) + \xi \delta N} \right)^2 + \theta \phi(J^\delta(x_0)) \leq \rho^2,
\end{equation}

where \( \theta \) is defined in (39). Then for all \( k \leq N \), the iterates \( x^\delta_k \) are in \( B_\rho(x^*) \), and we have the estimate (40) for \( k = 0, \ldots, N \).

**Proof.** We use induction over \( k \leq N \). For \( k = 0 \), \( x_0^\delta \) is in \( B_\rho(x^*) \) by (42). Let \( x_l^\delta \in B_\rho(x^*) \) for \( l = 0, \ldots, k, k < N \). We show that \( x_{k+1}^\delta \in B_\rho(x^*) \).
From (41) (with the sum up to the index $k - 1$) and (40), we may estimate

$$
\|e_{k+1}^\delta\|^2 \leq \left( \sqrt{\|e_k^\delta\|^2 + \frac{\theta}{(1 - L \delta)} J^\delta(x_0) + \xi \delta k} \right)^2 + \theta \|\nabla J_k^\delta\|^2.
$$

Using (13) for the last term on the right-hand side, by Lemma 3.2 and since $J^\delta(x_k) \leq J^\delta(x_0)$, we observe that $\|e_{k+1}^\delta\| \leq \rho$, thus $x_{k+1} \in B_\rho(x^*)$. Induction yields the assertion. The estimate (40) follows from Lemma 3.5.

Since it is well-known that the Landweber iteration has to be stopped for noisy data, we have to introduce a stopping criterion. Here we choose a simple a-priori rule: for each noise level $\delta$ define the stopping index $N_\delta$ such that

$$
\text{(43)} \quad \lim_{\delta \to 0} N_\delta = \infty, \quad \lim_{\delta \to 0} N_\delta \delta = 0, \quad (N_\delta + 1)\delta \leq \frac{\rho}{2\xi}.
$$

We have the following theorem.

**Theorem 3.6.** Let Assumption [I] and NC(0, $\beta$) hold for some $\beta \in \mathbb{R}$. Let $x_0$ be close to $x^*$ such that

$$
\text{(44)} \quad \|e_0\|^2 + \frac{2\theta}{(1 - L)} J(x_0) + \theta \phi(J(x_0)) \leq \frac{1}{16} \rho^2.
$$

Let $\{\delta_l\}$ be a sequence of noise levels associated to noisy data via (11), and let them be sufficiently small such that for all $l \in \mathbb{N}$

$$
\delta_l < \frac{1}{2} - \frac{L}{2}, \quad \phi(J(x_0) + \psi(\delta_l)) \leq \phi(J(x_0)) + \frac{\rho^2}{8\theta}, \quad \frac{2\theta}{(1 - L)} \psi(\delta_l) \leq \frac{\rho^2}{8}.
$$

holds. Let the stopping index be chosen as in (43). Then $x_{N_\delta_l}^\delta$ is in $B_\rho(x^*)$ and hence has a weakly convergent subsequence.

If $x \to \nabla J(x)$ is weakly sequentially closed on $B_\rho(x^*)$, then a limit of this subsequence is a stationary point of $J$. Assume additionally that $x^*$ is the unique stationary point of $J$ in $B_\rho(x^*)$. Then

$$
x_{N_\delta_l}^\delta \to x^*, \quad \text{as } \delta_l \to 0.
$$

**Proof.** Since $\delta_l$ is small, from (14) it follows that $J^\delta$ is Lipschitz with $L \delta < 1$ and $\frac{1}{1 - L \delta} \leq \frac{1}{1 - L}$ for all $\delta = \delta_l$. With (43) and (14), it may be verified that (12) holds for all $\delta_l$ and with $N = N_\delta_l + 1$. Thus, by Proposition 3 the iterates satisfy $x_k^{\delta_l} \in B_\rho(x^*)$ for all indices $k$ up to the stopping index $N_\delta_l + 1$. In particular, $x_{N_\delta_l}^{\delta_l}$ is bounded and has a weakly convergent subsequence.

Since $x_{N_\delta_l+1}^{\delta_l} \in B_\rho(x^*)$, we have by (35) that

$$
\|\nabla J_{N_\delta_l}^{\delta_l}\|^2 \leq \frac{2}{1 - L} \left( J_{N_\delta_l}^{\delta_l} - J_{N_\delta_l-1}^{\delta_l} \right).
$$

From (12) we find that

$$
\|\nabla J_{N_\delta_l}^{\delta_l}\|^2 \leq \frac{2}{1 - L} \left| J_{N_\delta_l} - J_{N_\delta_l-1} \right| + \frac{2}{1 - L} \psi(\delta_l).
$$

By Corollary 1 the sequence $x_k$ is in $B_\rho(x^*)$, hence by Lemma 3.1 the sequence $J_k$ is decreasing and hence convergent. Since $N_\delta_l \to \infty$, and $\delta_l \to 0$, we conclude by (11) that

$$
\lim_{\delta_l \to 0} \|\nabla J_{N_\delta_l}\| = 0.
$$
Then, by weakly closedness, 
\[ x_{N_{j_i}}^\delta \to \tilde{x} \quad \text{and} \quad \nabla J(x_{N_{j_i}}^\delta) \to 0 \implies \nabla J(\tilde{x}) = 0. \]

Hence the limit point is a stationary point. If the stationary point in \( B_\rho(x^*) \) is unique, then any subsequence has a weakly convergent subsequence with limit \( x^* \), thus \( x_{N_{j_i}}^\delta \) must converge weakly to \( x^* \). \( \square \)

4. Strong convergence. The next step in the analysis concerns a proof of strong convergence of the iterations. As could be expected, this requires additional conditions, namely the functional has to be \( \gamma \) balanced and satisfies NC(\( \gamma, \beta \)).

**Lemma 4.1.** Let Assumption 1 and NC(0, \( \beta \)) hold for some \( \beta \in \mathbb{R} \). Let \( x_0 - x^* \) small enough that (37) holds. Then

\[
\sum_{k=1}^{\infty} |\langle \nabla J_k, e_k \rangle| < \infty.
\]

**Proof.** From (32) we obtain that for any integer \( n_1 < n_2 \) that

\[
\sum_{k=n_1}^{n_2} |\langle \nabla J_k, e_k \rangle| = \sum_{k=n_1}^{n_2} |\langle e_{k+1} - e_k, e_k + e_{k+1} \rangle| \\
= -\sum_{k=n_1}^{n_2} [||e_{k+1}||^2 - ||e_k||^2] = -||e_{n_2+1}||^2 + ||e_{n_1}||^2.
\]

Moreover, from \( \langle \nabla J_k, e_{k+1} \rangle = \langle \nabla J_k, e_k \rangle - ||\nabla J_k||^2 \) we obtain

\[
2 \sum_{k=n_1}^{n_2} \langle \nabla J_k, e_k \rangle = -||e_{n_2+1}||^2 + ||e_{n_1}||^2 + \sum_{k=n_1}^{n_2} ||\nabla J_k||^2.
\]

We split the sum by defining the the index sets \( I_1 = \{ k \in [n_1, n_2] : \langle \nabla J_k, e_k \rangle \geq 0 \} \) and \( I_2 = \{ k \in [n_1, n_2] : \langle \nabla J_k, e_k \rangle < 0 \} \) and use (16) to find

\[
2 \sum_{k=n_1}^{n_2} |\langle \nabla J_k, e_k \rangle| = 2 \sum_{k \in I_1} \langle \nabla J_k, e_k \rangle - 2 \sum_{k \in I_2} \langle \nabla J_k, e_k \rangle \\
= -||e_{n_2+1}||^2 + ||e_{n_1}||^2 + \sum_{k=n_1}^{n_2} ||\nabla J_k||^2 - 4 \sum_{k \in I_2} \langle \nabla J_k, e_k \rangle \\
\leq -||e_{n_2+1}||^2 + ||e_{n_1}||^2 + (1 + 4\beta^+) \sum_{k=n_1}^{n_2} ||\nabla J_k||^2.
\]

According to Lemmas 3.3 and 3.1 the right-hand side is uniformly bounded. \( \square \)

**Lemma 4.2.** Let Assumption 1 hold. Suppose that \( J \) satisfies NC(\( \gamma, \beta \)) and is \( \gamma \)-balanced for some \( \gamma \geq 0 \) and for some \( \beta \in \mathbb{R} \). For a subsequence \( \{e_{k_m}\}_m \) assume that \( \liminf_m ||e_{k_m}|| \geq c_0 > 0 \). Then, with \( \tau \) and \( n_0 \) from (21), for any \( 0 \leq s < k_m, k_m \geq n_0 \),

\[
-\langle \nabla J_s, e_{k_m} \rangle \leq \frac{1}{\tau} \langle \nabla J_s, e_s \rangle + \frac{\beta}{\tau} ||\nabla J(x_s)||^2.
\]

**Proof.** By (21), we find a \( \tau > 0 \) with

\[
J(x^* - \tau e_{k_m}) \leq \gamma J(x^* + e_{k_m}) = \gamma J(x_{k_m}) \quad \forall k_m \geq n_0
\]
Thus if $s \leq k_m$, we have by Lemma 5.1 that $J(x^s - \tau e_{k_m}) \leq \gamma J(x_{k_m}) \leq \gamma J(x_s)$. We apply (16) with $x_2 = x_s$ and $x_1 = x^s - \tau e_{k_m}$. It then holds that $x_1, x_2 \in B_p(x^s)$. This yields
\[
\beta \|\nabla J(x_s)\|^2 \geq \langle \nabla J(x_s), x^s - \tau e_{k_m} - x_s \rangle = -\langle \nabla J_s, e_s \rangle - \tau \langle \nabla J_s, e_{k_m} \rangle.
\]
\[
\Box
\]

Summing up we arrive at the following theorem about the convergence of the exact iteration.

**Theorem 4.3.** Let Assumption 1 hold and suppose that $J$ satisfies NC($\gamma, \beta$) and is $\gamma$-balanced for some $\gamma \geq 0$ and for some $\beta \in \mathbb{R}$. Let $x_0 - x^*$ be small enough such that (37) holds. Then $\{x_k\}_k$ has a strongly convergent subsequence with a limit that is a stationary point. If $x^*$ is the unique stationary point of $J$ in $B_p(x^*)$, then the sequence $\{x_k\}_k$ converges to $x^*$.

**Proof.** By Corollary 1, $e_k$ is bounded for all $k$. Hence, there exist a subsequence such that $\|e_k\|$ is convergent. If $e_k$ has a strongly convergent subsequence with limit 0, then the assertion follows trivially. Otherwise, for any subsequence we have $\liminf \|e_k\| \geq 0$, in particular, also for one for which $\|e_k\|$ is convergent. Take such a subsequence and write for $k_m > k_n \geq n_0$, where $n_0$ is the index in (21)
\[
\|e_{k_n} - e_{k_m}\|^2 = \|e_{k_n}\|^2 - \|e_{k_m}\|^2 + 2\langle e_{k_m} - e_{k_n}, e_{k_m} \rangle
\]
\[
= \|e_{k_n}\|^2 - \|e_{k_m}\|^2 - 2\sum_{s=k_n}^{k_m-1} \langle \nabla J_s, e_{k_m} \rangle.
\]
From Lemma 4.2 we obtain that
\[
\|e_{k_n} - e_{k_m}\|^2 \leq \|e_{k_n}\|^2 - \|e_{k_m}\|^2 + \frac{2}{\tau} \sum_{s=k_n}^{k_m-1} \|\nabla J_s, e_s\|^2 + \frac{2\beta}{\tau} \sum_{s=k_n}^{k_m-1} \|\nabla J(x_s)\|^2
\]
\[
\leq \|e_{k_n}\|^2 - \|e_{k_m}\|^2 + \frac{2}{\tau} \sum_{s=k_n}^{k_m-1} \|\nabla J_s, e_s\|^2 + \frac{2\beta}{\tau} \sum_{s=k_n}^{k_m-1} \|\nabla J(x_s)\|^2.
\]
By (16) and since $\|e_{k_m}\|$ is convergent, we may find for any given $\epsilon$ an $\bar{n}_0 \geq n_0$ such that for all $k_m > k_n > \bar{n}_0$ the right-hand side is smaller than $\epsilon$. Thus, $e_{k_m}$ is a Cauchy sequence and hence convergent. Since by Lemma 3.1, $\nabla J_{k_m} \to 0$, and $\nabla J$ is continuous, it follows that the limit $\tilde{x}$ must be a stationary point. If $x^*$ is the only possibility of such a limit, it follows by a standard subsequence argument, that $x_k$ must converge to $x^*$.

We now come to the main result of strong convergence in the noisy case. Concerning the stopping criterion, we define for each noise level $\delta$ the stopping index $N_\delta$ according to (43). Then we have the following theorem.

**Theorem 4.4.** Let Assumption 1 hold and suppose that $J$ satisfies NC($\gamma, \beta$) and is $\gamma$-balanced for some $\gamma \geq 0$ and for some $\beta \in \mathbb{R}$. Let $x_0 - x^*$ be small enough such that (37) holds. Let the sequence of noise levels $\delta_l \to 0$ be sufficiently small such that (45) holds, and let the stopping index be chosen as in (43).

Assume that $x^*$ is the unique stationary point of $J$ in $B_p(x^*)$. Then
\[
\lim_{\delta_l \to 0} x_{N_{\delta_l}} = x^*.
\]
Thus for $N_{\gamma}$, Corollary 2. Let Assumption 1 hold, let the level sets $\{J < \alpha\} \cap B_p(x^*)$ be convex, and $J$ be 1-balanced. With $x_0 - x^*$, $\delta$, and the stopping index as in Theorem 4.4 for $\gamma = 1$, and if $x^*$ is the unique stationary point of $J$ in $B_p(x^*)$, then we have
\[
\lim_{\delta \to 0} x_{N_{\gamma}}^\delta = x^*.
\]

5. Conclusion. In this paper, we considered gradient descent iterations for functionals with Lipschitz-continuous derivative. We introduced new restrictions on the nonlinearity of the problem, namely the conditions $NC(\gamma, \beta)$ and the $\gamma$-balancing conditions. We have shown that they generalize several classical nonlinearity conditions.
The first main result concern weak convergence for the exact and noisy case of gradient iterations if the condition $NC(0, \beta)$ with some $\beta \in \mathbb{R}$ holds and using an a-priori stopping rule.

Strong convergence is verified in the exact case if $NC(\gamma, \beta)$, $\beta \in \mathbb{R}$, $\gamma \in [0, 1]$, holds and the functional is $\gamma$-balanced. With a stopping rule and if $x^*$ is the unique global minimum, strong convergence in the noisy case is verified under the same conditions.

REFERENCES

[1] K. J. Arrow and A. C. Enthoven, Quasi-concave programming, *Econometrica*, 29 (1961), 779–800.
[2] M. Avriel, W. E. Diewert, S. Schaible and I. Zang, *Generalized Concavity* vol. 63 of Classics in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2010.
[3] A. Bakushinsky and A. Goncharsky, *Ill-posed Problems: Theory and Applications* vol. 301 of Mathematics and its Applications, Kluwer Academic Publishers Group, Dordrecht, 1994.
[4] A. B. Bakushinsky and M. Y. Kokurin, *Iterative Methods for Approximate Solution of Inverse Problems*, vol. 577 of Mathematics and Its Applications (New York), Springer, Dordrecht, 2004.
[5] A. Bakushinsky and A. Smirnova, A posteriori stopping rule for regularized fixed point iterations, *Nonlinear Anal.*, 64 (2006), 1255–1261.
[6] A. Bakushinsky and A. Smirnova, Iterative regularization and generalized discrepancy principle for monotone operator equations, *Numer. Funct. Anal. Optim.*, 28 (2007), 13–25.
[7] A. B. Bakushinsky, M. Y. Kokurin and A. Smirnova, *Iterative Methods for Ill-Posed Problems*, vol. 54 of Inverse and Ill-posed Problems Series, Walter de Gruyter GmbH & Co. KG, Berlin, 2011.
[8] F. E. Browder and W. V. Petryshyn, Construction of fixed points of nonlinear mappings in Hilbert space, *J. Math. Anal. Appl.*, 20 (1967), 197–228.
[9] A.-P. Calderón, On an inverse boundary value problem, in *Seminar on Numerical Analysis and its Applications to Continuum Physics* (Rio de Janeiro, 1980), Soc. Brasil. Mat., Rio de Janeiro, 1980, 65–73.
[10] H. W. Engl, M. Hanke and A. Neubauer, *Regularization of Inverse Problems* vol. 375 of Mathematics and its Applications, Kluwer, Dordrecht, 1996.
[11] M. Haltmeier, A. Leitão and O. Scherzer, Kaczmarz methods for regularizing nonlinear ill-posed equations. I. Convergence analysis, *Inverse Probl. Imaging*, 1 (2007), 289–298.
[12] M. Hanke, A regularizing Levenberg-Marquardt scheme, with applications to inverse groundwater filtration problems, *Inverse Problems*, 13 (1997), 79–95.
[13] M. Hanke, A. Neubauer and O. Scherzer, A convergence analysis of the Landweber iteration for nonlinear ill-posed problems, *Numer. Math.*, 72 (1995), 21–37.
[14] T. Hein and K. S. Kazimierski, Accelerated Landweber iteration in Banach spaces, *Inverse Problems*, 26 (2002), Art. 055002, 17 pages.
[15] N. S. Hoang and A. G. Ramm, Dynamical systems gradient method for solving nonlinear equations with monotone operators, *Acta Appl. Math.*, 106 (2009), 473–499.
[16] N. S. Hoang and A. G. Ramm, The dynamical systems method for solving nonlinear equations with monotone operators, *Asian-Eur. J. Math.*, 3 (2010), 57–105.
[17] Q. Jin, A general convergence analysis of some Newton-type methods for nonlinear inverse problems, *SIAM J. Numer. Anal.*, 49 (2011), 549–573.
[18] B. Kaltenbacher, Some Newton-type methods for the regularization of nonlinear ill-posed problems, *Inverse Problems*, 13 (1997), 729–753.
[19] B. Kaltenbacher, A. Neubauer and O. Scherzer, *Iterative Regularization Methods for Nonlinear Ill-Posed Problems* vol. 6 of Radon Series on Computational and Applied Mathematics, Walter de Gruyter GmbH & Co. KG, Berlin, 2008.
[20] A. Neubauer, On Landweber iteration for nonlinear ill-posed problems in Hilbert scales, *Numer. Math.*, 85 (2000), 309–328.
[21] A. Neubauer, Some generalizations for Landweber iteration for nonlinear ill-posed problems, *Journal Inverse and Ill-posed Problems*, 24 (2016), 393–406.
S. S. Pereverzyev, R. Pinnau and N. Siedow. Regularized fixed-point iterations for nonlinear inverse problems. *Inverse Problems*, **22** (2006), 1–22.

A. G. Ramm, Dynamical systems method for ill-posed equations with monotone operators. *Commun. Nonlinear Sci. Numer. Simul.*, **10** (2005), 935–940.

O. Scherzer, Convergence criteria of iterative methods based on Landweber iteration for solving nonlinear problems. *J. Math. Anal. Appl.*, **194** (1995), 911–933.

T. Schuster, B. Kaltenbacher, B. Hofmann and K. S. Kazimierski, *Regularization Methods in Banach Spaces* vol. 10 of Radon Series on Computational and Applied Mathematics, Walter de Gruyter GmbH & Co. KG, Berlin, 2012.

V. V. Vasin, Convergence of gradient-type methods for nonlinear equations, *Dokl. Akad. Nauk*, **359** (1998), 7–9.

Received June 2016; revised May 2017.

E-mail address: kindermann@indmath.uni-linz.ac.at