Convergence analysis of a two-point gradient method for nonlinear ill-posed problems

IOP Publishing Inverse Problems

Inverse Problems **33** (2017) 095004 (30pp) https://doi.org/10.1088/1361-6420/aa7ac7

Convergence analysis of a two-point gradient method for nonlinear ill-posed problems

Simon Hubmer[1] **and Ronny Ramlau**[2,3]

[1] Johannes Kepler University Linz, Doctoral Program Computational Mathematics, Altenbergerstraße 69, A-4040 Linz, Austria

[2] Johannes Kepler University Linz, Institute of Industrial Mathematics, Altenbergerstraße 69, A-4040 Linz, Austria

[3] Johann Radon Institute Linz, Altenbergerstraße 69, A-4040 Linz, Austria

E-mail: simon.hubmer@dk-compmath.jku.at, ronny.ramlau@jku.at and ronny.ramlau@ricam.oeaw.ac.at

Received 17 November 2016, revised 16 May 2017

Accepted for publication 21 June 2017

Published 1 August 2017

Abstract

We perform a convergence analysis of a two-point gradient method which is based on Landweber iteration and on Nesterov's acceleration scheme. Additionally, we show the usefulness of this method via two numerical example problems based on a nonlinear Hammerstein operator and on the nonlinear inverse problem of single photon emission computed tomography.

Keywords: two-point gradient method, Nesterov acceleration scheme, Landweber iteration, steepest descent, minimal error, regularization method, SPECT

(Some figures may appear in colour only in the online journal)

1. Introduction

In this paper, we deal with nonlinear inverse problems of the form

$$F(x) = y, \tag{1.1}$$

where $F : \mathcal{D}(F) \subset \mathcal{X} \to \mathcal{Y}$ is a continuously Fréchet-differentiable, nonlinear operator between real Hilbert spaces $\mathcal{X}$ and $\mathcal{Y}$. Throughout this paper we will assume that (1.1) has a solution x_*, which need not be unique. Furthermore, we assume that instead of y, we are only given noisy data y^δ satisfying

$$\left\| y - y^\delta \right\| \leqslant \delta. \tag{1.2}$$

Since we are interested in ill-posed problems, we need to use regularization methods in order to obtain stable approximations of solutions of (1.1). The two most prominent examples of such methods are *Tikhonov regularization* and *Landweber iteration*.

In *Tikhonov regularization*, one attempts to approximate an x_0-minimum-norm solution $x^\dagger$ of (1.1), i.e. a solution of $F(x) = y$ with minimal distance to a given initial guess x_0, by minimizing the functional

$$\mathcal{T}_\alpha^\delta(x) := \left\| F(x) - y^\delta \right\|^2 + \alpha \left\| x - x_0 \right\|^2 , \tag{1.3}$$

where α is a suitably chosen regularization parameter. Under very mild assumptions on F, it can be shown that the minimizers of $\mathcal{T}_\alpha^\delta$, usually denoted by x_α^δ, converge to $x^\dagger$ as $\delta \to 0$, given that α and the noise level δ are coupled in an appropriate way [6]. While for linear operators F the minimization of $\mathcal{T}_\alpha^\delta$ is straightforward, in the case of nonlinear operators F the computation of x_α^δ requires the global minimization of the then also nonlinear functional $\mathcal{T}_\alpha^\delta$, which is rather difficult and usually done using various iterative optimization algorithms.

This motivates the direct application of iterative algorithms for solving (1.1), the most popular of which being *Landweber iteration*, given by

$$\begin{aligned} x_{k+1}^\delta &= x_k^\delta + \omega F'(x_k^\delta)^*(y^\delta - F(x_k^\delta)), \\ x_0^\delta &= x_0, \end{aligned} \tag{1.4}$$

where ω is a scaling parameter and x_0 is again a given initial guess. If one uses the discrepancy principle, i.e. stops the iteration after k_* steps, where k_* is the smallest integer such that

$$\left\| y^\delta - F(x_{k_*}^\delta) \right\| \leqslant \tau\delta < \left\| y^\delta - F(x_k^\delta) \right\| , \qquad 0 \leqslant k < k_*, \tag{1.5}$$

with a suitable constant $\tau > 1$, then it was proven in [6] that under some additional assumptions, most notably the nonlinearity condition (2.1), Landweber iteration gives rise to a convergent regularization method.

One necessary assumption in the convergence analysis of Landweber iteration is that

$$\omega \left\| F'(x^\dagger) \right\|^2 \leqslant 1. \tag{1.6}$$

Although estimating a suitable value for ω is easy in the linear case, for example using the power method (see e.g. [6]), in the nonlinear case a good estimate is hard to obtain. The *steepest descent method* [25] overcomes this problem by using the following iteration

$$x_{k+1}^\delta = x_k^\delta + \alpha_k^\delta s_k^\delta, \qquad s_k^\delta = F'(x_k^\delta)^*(y^\delta - F(x_k^\delta)), \quad k \in \mathbb{N}_0, \tag{1.7}$$

with the iteration dependent stepsize α_k^δ defined via

$$\alpha_k^\delta := \frac{\left\| s_k^\delta \right\|^2}{\left\| F'(x_k^\delta) s_k^\delta \right\|^2}. \tag{1.8}$$

This has the advantage of not having to estimate a fixed scaling parameter ω at the cost of having to compute $F'(x_k^\delta)s_k^\delta$ at every iteration step. Another possibility of choosing an iteration dependent stepsize α_k^δ without having to estimate ω is given by the *minimal error method* [14], which will be considered in a later section.

As is well known [14], both Landweber iteration and the steepest descent/minimal error method are quite slow. Hence, acceleration strategies have to be used in order to speed them up to make them applicable in practise. Acceleration methods and their analysis for linear problems can be found for example in [6] and [7]. Unfortunately, since their convergence proofs are mainly based on spectral theory, their analysis cannot be generalized to nonlinear problems immediately. However, there are some acceleration strategies for Landweber iteration for nonlinear ill-posed problems, for example [17, 21].

As an alternative to (accelerated) Landweber-type methods, one could think of using second order iterative methods for solving (1.1), such as the Levenberg–Marquardt method [8, 11]

$$x_{k+1}^\delta = x_k^\delta + (F'(x_k^\delta)^* F'(x_k^\delta) + \alpha_k I)^{-1} F'(x_k^\delta)^*(y^\delta - F(x_k^\delta)), \tag{1.9}$$

or the iteratively regularized Gauss–Newton method [3, 13]

$$x_{k+1}^\delta = x_k^\delta + (F'(x_k^\delta)^* F'(x_k^\delta) + \alpha_k I)^{-1} (F'(x_k^\delta)^*(y^\delta - F(x_k^\delta)) + \alpha_k(x_0 - x_k^\delta)). \tag{1.10}$$

The advantage of those methods [14] is that they require much less iterations to meet their respective stopping criteria than for example Landweber iteration or the steepest descent method. However, each of those iterations might take considerably longer than one step of Landweber iteration, due to the fact that in both cases a linear system involving the operator

$$F'(x_k^\delta)^* F'(x_k^\delta) + \alpha_k I \tag{1.11}$$

has to be solved. In practical applications, this usually means that a huge linear system of equations has to be solved, which often proves to be costly, if not impossible. Hence, accelerated Landweber type methods avoiding this drawback are desirable in practise.

An accelerated gradient method which also works remarkably well for nonlinear, albeit convex and well-posed optimization problems of the form

$$\min\{\Phi(x) \,|\, x \in \mathcal{X}\} \tag{1.12}$$

was first introduced by Nesterov in [16] and is given by

$$\begin{aligned} z_k &= x_k + \frac{k-1}{k+\alpha-1}(x_k - x_{k-1}), \\ x_{k+1} &= z_k - \omega(\nabla\Phi(z_k)), \end{aligned} \tag{1.13}$$

where again ω is a given scaling parameter and $\alpha \geqslant 3$ (with $\alpha = 3$ being common practise). This so-called *Nesterov acceleration scheme* is of particular interest, since not only is it extremely easy to implement, but Nesterov himself was also able to prove that it generates a sequence of iterates x_k for which there holds $\|\Phi(x_k) - \Phi(x_*)\| = \mathcal{O}(k^{-2})$, where x_* is any solution of (1.12). This is a big improvement over the classical rate $\mathcal{O}(k^{-1})$. For $\alpha > 3$ the even further improved rate $\mathcal{O}(k^{-2})$ was recently proven in [1].

Combined with a projection step, Nesterov's acceleration scheme can also be used to solve convex (and even non-smooth) optimization problems of the form

$$\min\{\Phi(x) + \Psi(x) \,|\, x \in \mathcal{X}\}, \tag{1.14}$$

and as such serves as the basis of the highly successful FISTA algorithm [2] for the fast solution of linear ill-posed problems with sparsity constraints.

Even though for general nonlinear operators F it is non-convex, one could think of applying Nesterov's acceleration scheme to the functional

$$\Phi(x) := \frac{1}{2}\left\|F(x) - y^\delta\right\|^2, \tag{1.15}$$

which leads to the algorithm

$$\begin{aligned} z_k^\delta &= x_k^\delta + \frac{k-1}{k+\alpha-1}(x_k^\delta - x_{k-1}^\delta), \\ x_{k+1}^\delta &= z_k^\delta + \alpha_k^\delta F'(z_k^\delta)^*(y^\delta - F(z_k^\delta)), \\ x_0^\delta &= x_{-1}^\delta = x_0, \end{aligned} \tag{1.16}$$

which in this form was first proposed in [12] to accelerate Landweber iteration for solving (nonlinear) ill-posed problems. Although no convergence analysis for (1.16) could be given, the numerical examples presented in [12] clearly show its usefulness and acceleration effect. Motivated by this, a slightly modified version of (1.16) promoting sparsity was used [24] and one of the authors of that paper, Neubauer, went on to show that for linear operators F and combined with a suitable stopping rule, (1.16) gives rise to a convergent regularization method [20]. This serves as motivation for considering general iteration methods of the form

$$\begin{aligned} z_k^\delta &= x_k^\delta + \lambda_k^\delta(x_k^\delta - x_{k-1}^\delta), \\ x_{k+1}^\delta &= z_k^\delta + \alpha_k^\delta s_k^\delta, \qquad s_k^\delta := F'(z_k^\delta)^*(y^\delta - F(z_k^\delta)), \\ x_0^\delta &= x_{-1}^\delta = x_0, \end{aligned} \tag{1.17}$$

which, for further reference, we will call *two-point gradient* (TPG) *methods* in this paper, since they require the use of the previous two iterates at every iteration step. Following the usual convention, we will drop the superscript δ whenever the iteration (1.17) with exact data $y^\delta = y$, i.e. $\delta = 0$, is considered.

The subsequent paper is structured as follows: in the next section, i.e. section 2, we present a convergence analysis of general TPG methods of the form (1.17), based on the classical convergence analysis of gradient based iterative regularization methods (see [14, 25]). This analysis will require certain abstract conditions on λ_k^δ and α_k^δ, which we will show to be satisfied for the steepest descent and the minimal error stepsizes and suitable choices of λ_k^δ in section 3. Afterwards, we will test the resulting TPG methods on both a nonlinear Hammerstein operator and a nonlinear single-photon-emission computed tomography (SPECT) example problem, numerically showing a considerable acceleration effect. Finally, we summarize our findings in section 5, discussing the results and providing a short outlook.

2. Convergence analysis

For the analysis of TPG methods of the form (1.17), we will need a few assumptions which are quite similar to the assumptions needed for the analysis of Landweber iteration or the steepest descent method [25]. Firstly, we will need the following local nonlinearity condition:

$$\begin{aligned} &\|F(x) - F(\tilde{x}) - F'(x)(x - \tilde{x})\| \leqslant \eta \, \|F(x) - F(\tilde{x})\|, \qquad \eta < \frac{1}{2}, \\ &x, \tilde{x} \in \mathcal{B}_{4\rho}(x_0) \subset \mathcal{D}(F), \end{aligned} \tag{2.1}$$

where $\mathcal{B}_{4\rho}(x_0)$ denotes the closed ball around x_0 with radius 4ρ. Assuming this condition to hold will allow the application of the following.

Lemma 2.1. *Let $\rho, \varepsilon > 0$ be such that*

$$\|F(x) - F(\tilde{x}) - F'(x)(x - \tilde{x})\| \leqslant c(x, \tilde{x}) \|F(x) - F(\tilde{x})\|, \quad x, \tilde{x} \in \mathcal{B}_\rho(x_0) \subset \mathcal{D}(F), \tag{2.2}$$

where $c(x, \tilde{x}) \geqslant 0$ and $c(x, \tilde{x}) < 1$ if $\|x - \tilde{x}\| \leqslant \varepsilon$. If $F(x) = y$ is solvable in $\mathcal{B}_\rho(x_0)$, then a unique x_0-minimum-norm solution exists. It is characterized as the solution $x^\dagger$ of $F(x) = y$ in $\mathcal{B}_\rho(x_0)$ satisfying the condition

$$x^\dagger - x_0 \in \mathcal{N}(F'(x^\dagger))^\perp. \tag{2.3}$$

Proof. [14, proposition 2.1] □

It will be necessary to place some restrictions on the stepsizes α_k^δ and the combination parameters λ_k^δ. Minimal requirements on their values are:

$$\lambda_0^\delta = 0, \qquad 0 \leqslant \lambda_k^\delta \leqslant 1, \ \forall k \in \mathbb{N}, \qquad \alpha_k^\delta \geqslant 0, \ \forall k \in \mathbb{N}. \tag{2.4}$$

With this, we can prove the following important proposition.

Proposition 2.2. *Assume that (2.1) and (2.4) hold and that equation $F(x) = y$ has a solution x_* in $\mathcal{B}_\rho(x_0) = \mathcal{B}_\rho(x_{-1})$ and let $x_k^\delta, x_{k-1}^\delta \in \mathcal{B}_\rho(x_*)$. Let*

$$\left\|y^\delta - F(z_k^\delta)\right\| > \tau\delta, \tag{2.5}$$

with τ satisfying

$$\tau > 2\frac{1+\eta}{1-2\eta}. \tag{2.6}$$

Setting

$$\Delta_k := \left\|x_k^\delta - x_*\right\|^2 - \left\|x_{k-1}^\delta - x_*\right\|^2, \tag{2.7}$$

and

$$\Psi := (1 - 2\eta) - 2\tau^{-1}(1 + \eta) > 0, \tag{2.8}$$

there holds

$$\Delta_{k+1} \leqslant \lambda_k^\delta \Delta_k + \lambda_k^\delta(\lambda_k^\delta + 1)\left\|x_k^\delta - x_{k-1}^\delta\right\|^2 - (1 + \Psi)\alpha_k^\delta \left\|y^\delta - F(z_k^\delta)\right\|^2 + (\alpha_k^\delta)^2 \left\|s_k^\delta\right\|^2. \tag{2.9}$$

Proof. Since $x_k^\delta, x_{k-1}^\delta \in \mathcal{B}_\rho(x_*)$, using the triangle inequality and $x_* \in \mathcal{B}_\rho(x_0)$, we get that $x_k^\delta, x_{k-1}^\delta \in \mathcal{B}_{2\rho}(x_0)$. Together with $\lambda_k^\delta \leqslant 1$, this implies

$$\begin{aligned}\left\|z_k^\delta - x_0\right\| &\leqslant \left\|z_k^\delta - x_k^\delta\right\| + \left\|x_k^\delta - x_0\right\| = \lambda_k^\delta \left\|x_k^\delta - x_{k-1}^\delta\right\| + \left\|x_k^\delta - x_0\right\| \\ &\leqslant \lambda_k^\delta \left\|x_k^\delta - x_*\right\| + \lambda_k^\delta \left\|x_* - x_{k-1}^\delta\right\| + \left\|x_k^\delta - x_0\right\| \leqslant 2\lambda_k^\delta\rho + 2\rho \leqslant 4\rho\end{aligned} \tag{2.10}$$

which shows that $z_k^\delta \in \mathcal{B}_{4\rho}(x_0)$. Hence, we can apply (2.1), which leads to

$$
\begin{aligned}
&\left\|x_{k+1}^\delta - x_*\right\|^2 - \left\|z_k^\delta - x_*\right\|^2 = \left\|x_{k+1}^\delta - z_k^\delta + z_k^\delta - x_*\right\|^2 - \left\|z_k^\delta - x_*\right\|^2 \\
&\quad = 2\left\langle x_{k+1}^\delta - z_k^\delta, z_k^\delta - x_* \right\rangle + \left\|x_{k+1}^\delta - z_k^\delta\right\|^2 \\
&\quad \overset{(1.17)}{=} 2\,\alpha_k^\delta \left\langle y^\delta - F(z_k^\delta), F'(z_k^\delta)(z_k^\delta - x_*) \right\rangle + (\alpha_k^\delta)^2 \left\|s_k^\delta\right\|^2 \\
&\quad = 2\,\alpha_k^\delta \left\langle y^\delta - F(z_k^\delta), y^\delta - y \right\rangle + 2\,\alpha_k^\delta \left\langle y^\delta - F(z_k^\delta), F(z_k^\delta) - y^\delta \right\rangle \\
&\qquad + 2\,\alpha_k^\delta \left\langle y^\delta - F(z_k^\delta), y - F(z_k^\delta) + F'(z_k^\delta)(z_k^\delta - x_*) \right\rangle + (\alpha_k^\delta)^2 \left\|s_k^\delta\right\|^2 \\
&\quad \leqslant 2\,\alpha_k^\delta \left\|y^\delta - F(z_k^\delta)\right\| \delta - 2\,\alpha_k^\delta \left\|y^\delta - F(z_k^\delta)\right\|^2 + (\alpha_k^\delta)^2 \left\|s_k^\delta\right\|^2 \\
&\qquad + 2\,\alpha_k^\delta \left\|y^\delta - F(z_k^\delta)\right\| \left\|y - F(z_k^\delta) + F'(z_k^\delta)(z_k^\delta - x_*)\right\| \\
&\quad \overset{(2.1)}{\leqslant} 2\,\alpha_k^\delta \left\|y^\delta - F(z_k^\delta)\right\| \delta - 2\,\alpha_k^\delta \left\|y^\delta - F(z_k^\delta)\right\|^2 + (\alpha_k^\delta)^2 \left\|s_k^\delta\right\|^2 \\
&\qquad + 2\,\alpha_k^\delta \eta \left\|y^\delta - F(z_k^\delta)\right\| \left\|F(x_*) - F(z_k^\delta)\right\| \\
&\quad \leqslant 2\,\alpha_k^\delta \left\|y^\delta - F(z_k^\delta)\right\| \delta - 2\,\alpha_k^\delta \left\|y^\delta - F(z_k^\delta)\right\|^2 + (\alpha_k^\delta)^2 \left\|s_k^\delta\right\|^2 \\
&\qquad + 2\,\alpha_k^\delta \eta \left\|y^\delta - F(z_k^\delta)\right\| \left(\left\|F(z_k^\delta) - y^\delta\right\| + \delta\right) \\
&\quad = \alpha_k^\delta \left\|y^\delta - F(z_k^\delta)\right\| \left(2\delta(1+\eta) - (1-2\eta)\left\|y^\delta - F(z_k^\delta)\right\|\right) \\
&\qquad - \alpha_k^\delta \left(\left\|F(z_k^\delta) - y^\delta\right\|^2 - \alpha_k^\delta \left\|s_k^\delta\right\|^2\right) \\
&\quad \overset{(2.5)}{\leqslant} \alpha_k^\delta \left\|y^\delta - F(z_k^\delta)\right\|^2 \left(2\tau^{-1}(1+\eta) - (1-2\eta)\right) \\
&\qquad - \alpha_k^\delta \left(\left\|F(z_k^\delta) - y^\delta\right\|^2 - \alpha_k^\delta \left\|s_k^\delta\right\|^2\right)
\end{aligned}
\tag{2.11}
$$

Hence, using (2.8), we arrive at the estimate

$$
\left\|x_{k+1}^\delta - x_*\right\|^2 \leqslant \left\|z_k^\delta - x_*\right\|^2 - (1+\Psi)\alpha_k^\delta \left\|F(z_k^\delta) - y^\delta\right\|^2 + (\alpha_k^\delta)^2 \left\|s_k^\delta\right\|^2 . \tag{2.12}
$$

Now, using the above inequality, we get

$$
\begin{aligned}
\Delta_{k+1} &= \left\|x_{k+1}^\delta - x_*\right\|^2 - \left\|x_k^\delta - x_*\right\|^2 \\
&\overset{(2.12)}{\leqslant} \left\|z_k^\delta - x_*\right\|^2 - \left\|x_k^\delta - x_*\right\|^2 - (1+\Psi)\alpha_k^\delta \left\|F(z_k^\delta) - y^\delta\right\|^2 + (\alpha_k^\delta)^2 \left\|s_k^\delta\right\|^2 \\
&= 2\left\langle z_k^\delta - x_k^\delta, x_k^\delta - x_* \right\rangle + \left\|z_k^\delta - x_k^\delta\right\|^2 - (1+\Psi)\alpha_k^\delta \left\|F(z_k^\delta) - y^\delta\right\|^2 + (\alpha_k^\delta)^2 \left\|s_k^\delta\right\|^2 \\
&\overset{(1.17)}{=} -2\lambda_k^\delta \left\langle x_{k-1}^\delta - x_k^\delta, x_k^\delta - x_* \right\rangle + (\lambda_k^\delta)^2 \left\|x_k^\delta - x_{k-1}^\delta\right\|^2 \\
&\qquad\qquad - (1+\Psi)\alpha_k^\delta \left\|F(z_k^\delta) - y^\delta\right\|^2 + (\alpha_k^\delta)^2 \left\|s_k^\delta\right\|^2 \\
&= -\lambda_k^\delta \left(\left\|x_{k-1}^\delta - x_k^\delta + x_k^\delta - x_*\right\|^2 - \left\|x_k^\delta - x_*\right\|^2 - \left\|x_k^\delta - x_{k-1}^\delta\right\|^2\right) \\
&\qquad\qquad + (\lambda_k^\delta)^2 \left\|x_k^\delta - x_{k-1}^\delta\right\|^2 - (1+\Psi)\alpha_k^\delta \left\|F(z_k^\delta) - y^\delta\right\|^2 + (\alpha_k^\delta)^2 \left\|s_k^\delta\right\|^2 \\
&= -\lambda_k^\delta \left(\left\|x_{k-1}^\delta - x_*\right\|^2 - \left\|x_k^\delta - x_*\right\|^2\right) + \lambda_k^\delta(\lambda_k^\delta + 1)\left\|x_k^\delta - x_{k-1}^\delta\right\|^2 \\
&\qquad\qquad - (1+\Psi)\alpha_k^\delta \left\|F(z_k^\delta) - y^\delta\right\|^2 + (\alpha_k^\delta)^2 \left\|s_k^\delta\right\|^2 \\
&= \lambda_k^\delta \Delta_k + \lambda_k^\delta(\lambda_k^\delta + 1)\left\|x_k^\delta - x_{k-1}^\delta\right\|^2 - (1+\Psi)\alpha_k^\delta \left\|F(z_k^\delta) - y^\delta\right\|^2 + (\alpha_k^\delta)^2 \left\|s_k^\delta\right\|^2 ,
\end{aligned}
$$

which yields the assertion. □

In order to stop the iteration, we will use the *discrepancy principle* with respect to z_k^δ, i.e. we will stop the iteration after k_* iterations, where $k_* = k_*(\delta, y^\delta)$ is the smallest integer such that

$$\left\|y^\delta - F(z_{k_*}^\delta)\right\| \leqslant \tau\delta < \left\|y^\delta - F(z_k^\delta)\right\|, \qquad 0 \leqslant k < k_*, \tag{2.13}$$

and use $z_{k_*}^\delta$ as an approximation of $x^\dagger$. For the constant τ, as suggested by proposition 2.2, we will use the condition

$$\tau > 2\frac{1+\eta}{1-2\eta}. \tag{2.14}$$

In the convergence analysis of Landweber iteration, one uses the fact that $\Delta_{k+1} \leqslant 0$ for all $k < k_*$, i.e. that x_{k+1}^δ is a better approximation of x_* than x_k^δ as long as the discrepancy principle (1.5) is not yet satisfied. We would like our TPG methods to share this property. Hence, in view of (2.9), we will use the following *coupling condition*:

$$\lambda_k^\delta(\lambda_k^\delta + 1)\left\|x_k^\delta - x_{k-1}^\delta\right\|^2 - \left(1 + \frac{\Psi}{\mu}\right)\alpha_k^\delta\left\|F(z_k^\delta) - y^\delta\right\|^2 + (\alpha_k^\delta)^2\left\|s_k^\delta\right\|^2 \leqslant 0 \tag{2.15}$$

which has to hold for all $0 \leqslant k < k_*$ with k_* determined by (2.13) and where μ is a constant satisfying $\mu > 1$. This implies $\Delta_{k+1} \leqslant \lambda_k^\delta\,\Delta_k$ and therefore, in view of $\lambda_0^\delta = 0$ and $\lambda_k^\delta \geqslant 0$ for all k, we inductively get that $\Delta_{k+1} \leqslant 0$[4].

Condition (2.15) essentially yields restrictions on the parameters λ_k^δ and α_k^δ. As a result, one has to ask if there exist choices of λ_k^δ and α_k^δ such that (2.15) is satisfied. For all stepsizes α_k^δ considered below, we will see that there holds

$$\alpha_k^\delta\left\|s_k^\delta\right\|^2 \leqslant \left\|F(z_k^\delta) - y^\delta\right\|^2, \tag{2.16}$$

and hence, a sufficient condition for (2.15) to hold is given by

$$\lambda_k^\delta(\lambda_k^\delta + 1)\left\|x_k^\delta - x_{k-1}^\delta\right\|^2 \leqslant \frac{\Psi}{\mu}\alpha_k^\delta\left\|F(z_k^\delta) - y^\delta\right\|^2. \tag{2.17}$$

Obviously, $\lambda_k^\delta = 0$ satisfies this inequality, which corresponds to classical Landweber type iterations. In finding other admissible choices of λ_k^δ and α_k^δ, one has to be careful, since both α_k^δ and z_k^δ might depend on λ_k^δ. Even for constant stepsizes $\alpha_k^\delta = \omega$ one is left with

$$\lambda_k^\delta(\lambda_k^\delta + 1)\left\|x_k^\delta - x_{k-1}^\delta\right\|^2 \leqslant \frac{\Psi}{\mu}\omega\left\|F(z_k^\delta) - y^\delta\right\|^2, \tag{2.18}$$

where it is not immediately clear how to choose λ_k^δ such that this inequality is satisfied. From the discrepancy principle (2.13), one can derive the sufficient condition

$$\lambda_k^\delta(\lambda_k^\delta + 1)\left\|x_k^\delta - x_{k-1}^\delta\right\|^2 \leqslant \frac{\Psi}{\mu}\omega(\tau\delta)^2, \tag{2.19}$$

which leads to the choice

$$\lambda_k^\delta = \min\left\{-\frac{1}{2} + \sqrt{\frac{1}{4} + \frac{\Psi\omega(\tau\delta)^2}{\mu\left\|x_k^\delta - x_{k-1}^\delta\right\|^2}}, 1\right\}, \tag{2.20}$$

[4] This is not necessarily the case for the classical Nesterov acceleration scheme (1.16), for which convergence can therefore not be proven using the presented framework.

where the minimum with 1 is taken in order to guarantee $0 \leqslant \lambda_k^\delta \leqslant 1$. As the numerical examples presented in section 4 will show, this choice indeed leads to a speedup compared to classical Landweber iteration which, however, decreases as $\delta \to 0$, which could be expected, since for $\delta = 0$, we get $\lambda_k^\delta = \lambda_k^0 = 0$ and hence, we recover the classical Landweber iteration, known to be slow.

One possibility for finding a sequence λ_k^δ, based on a backtracking search procedure, which takes nonzero values also for $\delta = 0$, satisfies condition (2.15) and leads to a considerable acceleration effect will be presented in section 3.

We now continue the convergence analysis of the TPG methods (1.17) by deducing the following proposition based on proposition 2.2 and the coupling condition (2.15).

Proposition 2.3. *Assume that* (2.1) *and* (2.4) *hold and that equation* $F(x) = y$ *has a solution* x_* *in* $\mathcal{B}_\rho(x_0) = \mathcal{B}_\rho(x_{-1})$. *Let* $k_* = k(\delta, y^\delta)$ *be chosen according to the stopping rule* (2.13), (2.14) *and assume that* (2.15) *holds for all* $0 \leqslant k < k_*$. *Then* x_k^δ *as in* (1.17) *is well-defined and*

$$\left\|x_{k+1}^\delta - x_*\right\| \leqslant \left\|x_k^\delta - x_*\right\|, \qquad \forall (-1) \leqslant k < k_*. \tag{2.21}$$

Moreover, $x_k^\delta \in \mathcal{B}_\rho(x_*) \subset \mathcal{B}_{2\rho}(x_0)$ *for all* $(-1) \leqslant k \leqslant k_*$ *and*

$$\left(\min_{0 \leqslant k < k_*} \{\alpha_k^\delta\}\right) k_* (\tau\delta)^2 \leqslant \sum_{k=0}^{k_*-1} \alpha_k^\delta \left\|y^\delta - F(z_k^\delta)\right\|^2 \leqslant (\bar{\mu}\Psi)^{-1} \left\|x_0^\delta - x_*\right\|^2, \tag{2.22}$$

where $\bar{\mu} = (\mu - 1)/\mu > 0$.

Proof. From (2.9) it follows for $k = 0$ that

$$\Delta_1 \leqslant \lambda_0^\delta \Delta_0 + \lambda_0^\delta(\lambda_0^\delta + 1) \left\|x_0^\delta - x_{-1}^\delta\right\|^2 - (1 + \Psi)\alpha_0^\delta \left\|y^\delta - F(z_0^\delta)\right\|^2 + (\alpha_0^\delta)^2 \left\|s_0^\delta\right\|^2.$$

Using (2.15) and $\lambda_0^\delta = 0$, we can deduce that

$$\begin{aligned} \Delta_1 &\leqslant \lambda_0^\delta(\lambda_0^\delta + 1) \left\|x_0^\delta - x_{-1}^\delta\right\|^2 - (1 + \Psi)\alpha_0^\delta \left\|y^\delta - F(z_0^\delta)\right\|^2 + (\alpha_0^\delta)^2 \left\|s_0^\delta\right\|^2 \\ &\overset{(2.15)}{\leqslant} -\frac{\mu - 1}{\mu} \Psi \alpha_0^\delta \left\|y^\delta - F(z_0^\delta)\right\|^2 = -\bar{\mu}\Psi\alpha_0^\delta \left\|y^\delta - F(z_0^\delta)\right\|^2 \leqslant 0, \end{aligned} \tag{2.23}$$

from which we get that $x_1^\delta \in \mathcal{B}_\rho(x_*)$. Now, we proceed inductively to show that

$$\Delta_{k+1} \leqslant -\bar{\mu}\Psi\alpha_k^\delta \left\|y^\delta - F(z_k^\delta)\right\|^2 \leqslant 0, \tag{2.24}$$

and $x_{k+1}^\delta \in \mathcal{B}_\rho(x_*)$ for all $0 \leqslant k < k_*$. To do so, we assume that this holds for all $0 \leqslant m \leqslant k$. Again using (2.9), we deduce that

$$\Delta_{k+1} \leqslant \lambda_k^\delta \Delta_k + \lambda_k^\delta(\lambda_k^\delta + 1) \left\|x_k^\delta - x_{k-1}^\delta\right\|^2 - (1 + \Psi)\alpha_k^\delta \left\|F(z_k^\delta) - y^\delta\right\|^2 + (\alpha_k^\delta)^2 \left\|s_k^\delta\right\|^2, \tag{2.25}$$

which, together with (2.15) and the induction hypothesis yields (2.24). From this, we can deduce $x_{k+1}^\delta \in \mathcal{B}_\rho(x_*) \subset \mathcal{B}_{2\rho}(x_0)$, which completes the induction.

Furthermore, from (2.24) we can deduce that

$$\bar{\mu}\Psi\alpha_k^\delta \left\|y^\delta - F(z_k^\delta)\right\|^2 \leqslant \left\|x_k^\delta - x_*\right\|^2 - \left\|x_{k+1}^\delta - x_*\right\|^2, \tag{2.26}$$

and hence, also

$$\sum_{k=0}^{k_*-1} \bar{\mu}\Psi\alpha_k^\delta \left\|y^\delta - F(z_k^\delta)\right\|^2 \leqslant \left\|x_0^\delta - x_*\right\|^2 - \left\|x_{k_*}^\delta - x_*\right\|^2 \leqslant \left\|x_0^\delta - x_*\right\|^2. \tag{2.27}$$

From this, we get the estimate

$$\left(\min_{0\leqslant k<k_*}\{\alpha_k^\delta\}\right) k_*(\tau\delta)^2 \leqslant \sum_{k=0}^{k_*-1} \alpha_k^\delta \left\|y^\delta - F(z_k^\delta)\right\|^2 \leqslant (\bar{\mu}\Psi)^{-1}\left\|x_0^\delta - x_*\right\|^2, \tag{2.28}$$

which yields the assertion. □

From the above proposition, we get the following simple corollary.

Corollary 2.4. *Under the assumptions of proposition 2.3, we have*

$$k_* \leqslant \left(\min_{0\leqslant k<k_*}\{\alpha_k^\delta\}\right)^{-1} \frac{\left\|x_0^\delta - x_*\right\|^2}{\bar{\mu}\Psi(\tau\delta)^2}. \tag{2.29}$$

If we are given exact data $y^\delta = y$, i.e. if $\delta = 0$, then (2.22) implies

$$\sum_{k=0}^{\infty} \alpha_k \left\|y - F(z_k)\right\|^2 < \infty, \tag{2.30}$$

as in this case $k_* = \infty$. Note that this only holds if $F(z_k) \neq y$ for all $k \in \mathbb{N}$, since otherwise the sum terminates in a finite number of steps. However, this is not a restriction, since if $F(z_k) = y$ for some k, then a solution is found and the iteration is terminated.

Combining (2.30) together with (2.15), we furthermore get that

$$\sum_{k=0}^{\infty} \lambda_k^0(\lambda_k^0 + 1)\left\|x_k - x_{k-1}\right\|^2 < \infty, \tag{2.31}$$

and

$$\sum_{k=0}^{\infty} (\alpha_k)^2 \left\|s_k\right\|^2 < \infty, \tag{2.32}$$

from which there obviously follows

$$\begin{aligned} &\lim_{k\to\infty} \alpha_k \left\|y - F(z_k)\right\|^2 = 0, \\ &\lim_{k\to\infty} \lambda_k^0(\lambda_k^0 + 1)\left\|x_k - x_{k-1}\right\|^2 = 0, \\ &\lim_{k\to\infty} (\alpha_k)^2 \left\|s_k\right\|^2 = 0. \end{aligned} \tag{2.33}$$

If, additionally, α_k^δ is bounded from below, i.e.

$$0 < \alpha_{\min}^\delta := \min_{k\in\mathbb{N}}\{\alpha_k^\delta\}, \tag{2.34}$$

then it even follows that

$$\lim_{k\to\infty} \left\|y - F(z_k)\right\| = 0. \tag{2.35}$$

If we can show that z_k converges as well, then we get convergence of the iteration to a solution of $F(x) = y$. In order to do this, we first have to show a couple of intermediate results. We start by showing that under certain assumptions, the sequence $\|z_k - x_*\|$ has a finite limit as $k \to \infty$.

Proposition 2.5. *Let x_* be a solution of $F(x) = y$, and let x_k be the iterates (1.17) with exact data, i.e. $\delta = 0$. Assume that $\|x_k - x_*\| \to \varepsilon$ as $k \to \infty$, where $\varepsilon \geqslant 0$ is a constant. If $\lambda_k^0 \|x_k - x_{k-1}\| \to 0$ and $\alpha_k \|s_k\| \to 0$ as $k \to \infty$, then there holds*

$$\lim_{k\to\infty} \|z_k - x_*\| = \varepsilon. \tag{2.36}$$

Proof. From the definition of the iterates (1.17), we have the inequality

$$\|z_k - x_*\| = \left\|x_k - x_* + \lambda_k^0(x_k - x_{k-1})\right\| \leqslant \|x_k - x_*\| + \lambda_k^0 \|x_k - x_{k-1}\| \tag{2.37}$$

and

$$\|x_{k+1} - x_*\| = \|z_k - x_* + \alpha_k s_k\| \leqslant \|z_k - x_*\| + \alpha_k \|s_k\|, \tag{2.38}$$

from which there follows

$$\|x_{k+1} - x_*\| - \alpha_k \|s_k\| \leqslant \|z_k - x_*\| \leqslant \|x_k - x_*\| + \lambda_k^0 \|x_k - x_{k-1}\|. \tag{2.39}$$

Taking the limit as $k \to \infty$ now yields the assertion. □

The following characterisation of the iterates x_k^δ will be useful later on.

Lemma 2.6. *For the iterates of the TPG methods (1.17) there holds*

$$x_k^\delta = x_0 + \sum_{i=0}^{k-1} \lambda_i^\delta (x_i^\delta - x_{i-1}^\delta) + \sum_{i=0}^{k-1} \alpha_i^\delta s_i^\delta, \tag{2.40}$$

as well as

$$x_l^\delta - x_j^\delta = \sum_{i=j}^{l-1} \lambda_i^\delta (x_i^\delta - x_{i-1}^\delta) + \sum_{i=j}^{l-1} \alpha_i^\delta s_i^\delta, \tag{2.41}$$

and

$$x_i^\delta - x_{i-1}^\delta = \sum_{m=0}^{i-2} \left(\prod_{n=m+1}^{i-1} \lambda_n^\delta \right) \alpha_m^\delta s_m^\delta + \alpha_{i-1}^\delta s_{i-1}^\delta. \tag{2.42}$$

Proof. The first two of the above statements follow immediately from (1.17). Hence, it remains to prove (2.42), which we do by induction. For $i = 1$ the statement follows immediately from (1.17). Assuming now that (2.42) holds for all $1 \leqslant l \leqslant i$, we get

$$\begin{aligned} x_{i+1}^\delta - x_i^\delta &\overset{(1.17)}{=} \lambda_i^\delta (x_i^\delta - x_{i-1}^\delta) + \alpha_i^\delta s_i^\delta \\ &= \lambda_i^\delta \left(\sum_{m=0}^{i-2} \left(\prod_{n=m+1}^{i-1} \lambda_n^\delta \right) \alpha_m^\delta s_m^\delta + \alpha_{i-1}^\delta s_{i-1}^\delta \right) + \alpha_i^\delta s_i^\delta \\ &= \sum_{m=0}^{i-1} \left(\prod_{n=m+1}^{i} \lambda_n^\delta \right) \alpha_m^\delta s_m^\delta + \alpha_i^\delta s_i^\delta, \end{aligned} \tag{2.43}$$

which concludes the induction and hence the lemma is shown. □

Lemma 2.7. *Assume that (2.1) holds, let $x_* \in \mathcal{B}_{4\rho}(x_0)$ be a solution of $F(x) = y$ and let $x_1, x_2 \in \mathcal{B}_{4\rho}(x_0)$. Then there holds*

$$\|F'(x_1)(x_* - x_2)\| \leqslant 2(1+\eta)\,\|F(x_1) - y\| + (1+\eta)\,\|F(x_2) - y\|\,. \tag{2.44}$$

Proof. *The proof of this lemma was already done in [25] and is re-stated here for the sake of completeness. Using (2.1), it follows that*

$$\begin{aligned}\|F'(x_1)(x_* - x_2)\| &= \|F'(x_1)(x_* - x_1 + x_1 - x_2)\| \\ &\leqslant \|-F(x_*) + F(x_1) + F'(x_1)(x_* - x_1) - F(x_1) + F(x_*)\| \\ &\quad + \|F(x_2) - F(x_1) + F'(x_1)(x_1 - x_2) - F(x_2) + F(x_1)\| \\ &\leqslant (1+\eta)\,\|F(x_1) - y\| + (1+\eta)\,\|F(x_1) - F(x_2)\| \\ &\leqslant 2(1+\eta)\,\|F(x_1) - y\| + (1+\eta)\,\|F(x_2) - y\|\,,\end{aligned} \tag{2.45}$$

which yields the assertion. □

In order to prove convergence in the case of exact data in theorem 2.8 below, we need the following additional assumption on the combination parameters λ_k^0:

$$\sum_{k=0}^{\infty} \lambda_k^0\,\|x_k - x_{k-1}\| < \infty. \tag{2.46}$$

Since under the previous assumptions $\|x_k - x_{k-1}\|$ can be bounded (by 2ρ), it follows that a sufficient condition for (2.46) to hold is given by

$$\sum_{k=0}^{\infty} \lambda_k^0 < \infty. \tag{2.47}$$

For λ_k^δ defined via (2.20), condition (2.47) is obviously satisfied. However, it is quite a restrictive condition, since it implies $\lambda_k^0 \to 0$ as $k \to \infty$. Comparing this with the classical Nesterov combination parameters $\lambda_k^\delta = (k-1)/(k+\alpha-1)$, which tend to 1 as $k \to \infty$ even for $\delta = 0$, we see that in order to achieve a non-negligible acceleration effect also for $\delta = 0$, one has to work with condition (2.46) instead of only the sufficient condition (2.47). In section 3, we will present an algorithm for choosing λ_k^δ such that (2.46) is satisfied and the numerical examples presented in section 4 will show that for this sequence, under a suitable choice of parameters, there holds $\lambda_k^\delta \to 1$ as $k \to \infty$, leading to the desired acceleration effect. Using (2.46), we can now prove the following.

Theorem 2.8. *Assume that (2.1) holds and that equation $F(x) = y$ has a solution x_* in $\mathcal{B}_\rho(x_0) = \mathcal{B}_\rho(x_{-1})$. Let $k_* = k_*(0, y) = \infty$, λ_k^δ and α_k^δ satisfy (2.4), (2.34) and (2.46) and assume that (2.15) holds for all $k \in \mathbb{N}$. Then the iterates z_k defined as in (1.17) with exact data $y^\delta = y$ converge to a solution of $F(x) = y$. If $\mathcal{N}(F'(x^\dagger)) \subset \mathcal{N}(F'(x))$ for all $x \in \mathcal{B}_{4\rho}(x^\dagger)$, then z_k converges to $x^\dagger$ as $k \to \infty$.*

Proof. This proof closely follows the corresponding proof for Landweber iteration given in [6]. Let x_* be a solution of $F(x) = y$ in $\mathcal{B}_\rho(x_0)$ and define

$$e_k := z_k - x_*. \tag{2.48}$$

From proposition 2.3 it follows that $\|x_k - x_*\|$ converges to some $\varepsilon \geqslant 0$ and hence, using (2.33) and proposition 2.5, we can deduce that $\|e_k\|$ converges to this same ε as well. We are now going to show that e_k is a Cauchy sequence. Given $j \geqslant k$, we choose some integer l between k and j with

$$\|y - F(z_l)\| \leqslant \|y - F(z_i)\|\,, \qquad \forall\, k \leqslant i \leqslant j. \tag{2.49}$$

We have

$$\|e_j - e_k\| \leqslant \|e_j - e_l\| + \|e_l - e_k\|\,, \tag{2.50}$$

and

$$\begin{aligned} \|e_j - e_l\|^2 &= 2\,\langle\, e_l - e_j, e_l \,\rangle + \|e_j\|^2 - \|e_l\|^2\,, \\ \|e_l - e_k\|^2 &= 2\,\langle\, e_l - e_k, e_l \,\rangle + \|e_k\|^2 - \|e_l\|^2\,. \end{aligned} \tag{2.51}$$

For $k \to \infty$, the last two terms on each of the right hand sides of the above equations converge to $\varepsilon^2 - \varepsilon^2 = 0$. We now show that $\langle\, e_l - e_k, e_l \,\rangle$ and $\langle\, e_l - e_j, e_l \,\rangle$ also tend to 0 as $k \to \infty$. For this we first consider:

$$\begin{aligned} |\langle\, e_l - e_k, e_l \,\rangle| &= |\langle\, z_l - z_k, e_l \,\rangle| = \left|\left\langle\, x_l - x_k + \lambda_l^0(x_l - x_{l-1}) - \lambda_k^0(x_k - x_{k-1}), e_l \,\right\rangle\right| \\ &\leqslant |\langle\, x_l - x_k, e_l \,\rangle| + \lambda_l^0\,|\langle\, x_l - x_{l-1}, e_l \,\rangle| + \lambda_k^0\,|\langle\, x_k - x_{k-1}, e_l \,\rangle| \\ &\leqslant |\langle\, x_l - x_k, e_l \,\rangle| + \lambda_l^0\,\|x_l - x_{l-1}\|\,\|e_l\| + \lambda_k^0\,\|x_k - x_{k-1}\|\,\|e_l\|\,. \end{aligned} \tag{2.52}$$

Now, using (2.33) and the fact that $\|e_k\|$ converges to ε, we get that

$$\lim_{k\to\infty} \left(\lambda_l^0\,\|x_l - x_{l-1}\|\,\|e_l\| + \lambda_k^0\,\|x_k - x_{k-1}\|\,\|e_l\|\right) = 0. \tag{2.53}$$

Hence, it remains to consider

$$\begin{aligned} |\langle\, x_l - x_k, e_l \,\rangle| &\overset{(2.41)}{=} \left|\left\langle \sum_{i=k}^{l-1} \lambda_i^0(x_i - x_{i-1}) + \sum_{i=k}^{l-1} \alpha_i s_i, e_l \right\rangle\right| \\ &\leqslant \sum_{i=k}^{l-1} \lambda_i^0\,|\langle\, x_i - x_{i-1}, e_l \,\rangle| + \sum_{i=k}^{l-1} \alpha_i\,|\langle\, s_i, e_l \,\rangle|\,. \end{aligned} \tag{2.54}$$

We now consider the above two sums separately, starting with the second one. By lemma 2.7, we have

$$\begin{aligned} \sum_{i=k}^{l-1} \alpha_i\,|\langle\, s_i, e_l \,\rangle| &= \sum_{i=k}^{l-1} \alpha_i\,|\langle\, y - F(z_i), F'(z_i)(z_l - x_*) \,\rangle| \\ &\leqslant \sum_{i=k}^{l-1} \alpha_i\,\|y - F(z_i)\|\,\|F'(z_i)(z_l - x_*)\| \\ &\overset{(2.44)}{\leqslant} 2(1+\eta) \sum_{i=k}^{l-1} \alpha_i\,\|y - F(z_i)\|^2 + (1+\eta) \sum_{i=k}^{l-1} \alpha_i\,\|y - F(z_i)\|\,\|y - F(z_l)\| \\ &\leqslant 3(1+\eta) \sum_{i=k}^{l-1} \alpha_i\,\|y - F(z_i)\|^2 \leqslant 3(1+\eta) \sum_{i=k}^{\infty} \alpha_i\,\|y - F(z_i)\|^2\,, \end{aligned} \tag{2.55}$$

where we have used (2.49). From this, it follows by using (2.30) that

$$\lim_{k\to\infty}\left(\sum_{i=k}^{l-1}\alpha_i\,|\langle\, s_i, e_l\,\rangle|\right)=0. \tag{2.56}$$

Next we consider

$$\sum_{i=k}^{l-1}\lambda_i^0\,|\langle\, x_i-x_{i-1}, e_l\,\rangle|\leqslant\sum_{i=k}^{l-1}\lambda_i^0\,\|x_i-x_{i-1}\|\,\|e_l\|\leqslant\sum_{i=k}^{\infty}\lambda_i^0\,\|x_i-x_{i-1}\|\,\|e_l\|\,. \tag{2.57}$$

Since $\|e_l\|$ is bounded, it immediately follows from (2.46) that

$$\lim_{k\to\infty}\left(\sum_{i=k}^{l-1}\lambda_i^0\,|\langle\, x_i-x_{i-1}, e_l\,\rangle|\right)=0. \tag{2.58}$$

Combining the above estimates, we arrive at $|\langle\, x_l-x_k, e_l\,\rangle|\to 0$, from which there follows that $|\langle\, e_l-e_k, e_l\,\rangle|\to 0$ as $k\to\infty$. Since it can similarly be shown that $|\langle\, e_l-e_j, e_l\,\rangle|\to 0$ as $k\to\infty$, it follows that

$$\lim_{k\to\infty}\|e_j-e_k\|=0, \tag{2.59}$$

from which we deduce that e_k and hence, also z_k is a Cauchy sequence and therefore convergent in the Hilbert space $\mathcal{X}$. Since $\|F(z_k)-y\|$ converges to 0, the limit of z_k is a solution of $F(x)=y$.

Now we turn to the second part of the proof. If $\mathcal{N}(F'(x^\dagger))\subset\mathcal{N}(F'(x))$ for all $x\in\mathcal{B}_{4\rho}(x^\dagger)$, then by the definition of the iterates (1.17) we have

$$\begin{aligned} z_{k+1}-z_k &= x_{k+1}+\lambda_{k+1}^0(x_{k+1}-x_k)-z_k=\alpha_k s_k+\lambda_{k+1}^0(x_{k+1}-x_k)\\ &=(1+\lambda_{k+1}^0)\alpha_k s_k+\lambda_{k+1}^0(z_k-x_k)=(1+\lambda_{k+1}^0)\alpha_k s_k+\lambda_{k+1}^0\lambda_k^0(x_k-x_{k-1}) \end{aligned}$$

and therefore

$$z_k-z_0=\sum_{i=0}^{k-1}(z_{i+1}-z_i)=\sum_{i=0}^{k-1}\left((1+\lambda_{i+1}^0)\alpha_i s_i+\lambda_{i+1}^0\lambda_i^0(x_i-x_{i-1})\right). \tag{2.60}$$

Since obviously $(1+\lambda_{i+1}^0)\alpha_i s_i\in\mathcal{R}(F'(z_i)^*)$ and since

$$\mathcal{R}(F'(z_i)^*)\subset\mathcal{N}(F'(z_i))^\perp\subset\mathcal{N}(F'(x^\dagger))^\perp\quad\text{for all } i\in\mathbb{N}, \tag{2.61}$$

it follows that

$$\sum_{i=0}^{k-1}(1+\lambda_{i+1}^0)\alpha_i s_i\in\mathcal{N}(F'(x^\dagger))^\perp. \tag{2.62}$$

Similarly as above, it can be seen via using lemma 2.6 that also

$$\sum_{i=0}^{k-1}\lambda_{i+1}^0\lambda_i^0(x_i-x_{i-1})\in\mathcal{N}(F'(x^\dagger))^\perp, \tag{2.63}$$

and we therefore conclude that

$$z_k - z_0 \in \mathcal{N}(F'(x^\dagger))^\perp \quad \text{for all } k \in \mathbb{N}. \tag{2.64}$$

Since this also holds for the limit of z_k and since $x^\dagger$ is the unique solution for which this condition holds (see lemma 2.1), this proves that $z_k \to x^\dagger$ as $k \to \infty$. □

In the next corollary, we deduce the convergence of x_k given the convergence of z_k.

Corollary 2.9. *Under the assumptions of theorem 2.8, we get that x_k converges to x_*, where x_* is the limit of z_k as $k \to \infty$.*

Proof. The statement follows immediately from

$$\|x_{k+1} - x_*\| \leqslant \|z_k - x_*\| + \alpha_k \|s_k\|, \tag{2.65}$$

together with (2.33). □

Next, we show that using the discrepancy principle (2.13) as a stopping rule, our TPG method (1.17) becomes a convergent regularization method, if we additionally assume that λ_k^δ depends continuously on δ for $\delta \to 0$.

Theorem 2.10. *Assume that (2.1) holds and that equation $F(x) = y$ has a solution x_* in $\mathcal{B}_\rho(x_0) = \mathcal{B}_\rho(x_{-1})$. Let $k_* = k_*(\delta, y^\delta)$ be chosen according to the discrepancy principle (2.13), (2.14) and assume that (2.15) holds for all $0 \leqslant k < k_*$. Assume that λ_k^δ and α_k^δ satisfy (2.4), (2.34) and (2.46) and that $\lambda_k^\delta \to \lambda_k^0$ as $\delta \to 0$. Then the iterates $z_{k_*}^\delta$ defined via (1.17) converge to a solution of $F(x) = y$, as $\delta \to 0$. If $\mathcal{N}(F'(x^\dagger)) \subset \mathcal{N}(F'(x))$ for all $x \in \mathcal{B}_{4\rho}(x^\dagger)$, then $z_{k_*}^\delta$ converges to $x^\dagger$ as $\delta \to 0$.*

Proof. Again this proof closely follows the corresponding proof for Landweber iteration given in [6]. Let x_* be the limit point of z_k (and hence, by corollary 2.9, also of x_k) given exact data y and let δ_n be a sequence converging to 0 as $n \to \infty$. Let furthermore $y_n := y^{\delta_n}$ be a sequence of noisy data with $\|y - y_n\| \leqslant \delta_n$ and let $k_n := k_*(\delta_n, y_n)$ be the stopping index determined via the discrepancy principle applied to the pair (δ_n, y_n). There are two cases. First, assume that k is a finite accumulation point of k_n. Without loss of generality, we can assume that $k_n = k$ for all $n \in \mathbb{N}$. Thus, from the definition of the discrepancy principle, it follows that

$$\left\| y_n - F(z_k^{\delta_n}) \right\| \leqslant \tau \delta_n. \tag{2.66}$$

As k is fixed, z_k^δ depends continuously on the data y^δ and we can take the limit $n \to \infty$ in the above inequality, which yields

$$z_k^{\delta_n} \to z_k, \quad F(z_k^{\delta_n}) \to F(z_k) = y, \text{ as } n \to \infty. \tag{2.67}$$

In other words, the kth iterate of Landweber iteration with exact data is a solution of $F(x) = y$ and hence, the iteration terminates with $z_k = x_*$, and $z_{k_n}^{\delta_n} \to x_*$ for this subsequence as $\delta_n \to 0$.

For the second case, assume that $k_n \to \infty$ as $n \to \infty$. For some k and $k_n > k + 1$, proposition 2.3 and $0 \leqslant \lambda_k^\delta \leqslant 1$ yield

$$\begin{aligned}
\left\| z_{k_n}^{\delta_n} - x_* \right\| &\leqslant \left\| x_{k_n}^{\delta_n} - x_* \right\| + \lambda_k^\delta \left\| x_{k_n}^{\delta_n} - x_* \right\| + \lambda_k^\delta \left\| x_{k_n-1}^{\delta_n} - x_* \right\| \\
&\leqslant \left\| x_k^{\delta_n} - x_* \right\| + \lambda_k^\delta \left\| x_k^{\delta_n} - x_* \right\| + \lambda_k^\delta \left\| x_k^{\delta_n} - x_* \right\| \\
&\leqslant 3 \left\| x_k^{\delta_n} - x_* \right\| \leqslant 3 \left\| x_k^{\delta_n} - x_k \right\| + 3 \left\| x_k - x_* \right\|.
\end{aligned} \tag{2.68}$$

If we fix some $\varepsilon > 0$, it follows from proposition 2.2 and from corollary 2.9 that we can fix some $k = k(\varepsilon)$ such that $\|x_k - x_*\| \leqslant \varepsilon/6$. Since, for fixed k, the iterates depend continuously on the data, there is an $n = n(\varepsilon, k)$ such that $\left\|x_k^{\delta_n} - x_k\right\| \leqslant \varepsilon/6$ for all $n > n(\varepsilon, k)$. Thus if we choose n sufficiently large, such that also $k_n > k + 1$, we get that

$$\left\|z_{k_n}^{\delta_n} - x_*\right\| \leqslant 3\left\|x_k^{\delta_n} - x_*\right\| \leqslant 3\left\|x_k^{\delta_n} - x_k\right\| + 3\left\|x_k - x_*\right\| \leqslant 3\,\frac{\varepsilon}{6} + 3\,\frac{\varepsilon}{6} = \varepsilon, \tag{2.69}$$

and therefore $z_{k_n}^{\delta_n} \to x_*$ as $n \to \infty$, which shows the first part of the assertion. If $\mathcal{N}(F'(x^\dagger)) \subset \mathcal{N}(F'(x))$ for all $x \in \mathcal{B}_{4\rho}(x^\dagger)$, then x_* can be chosen as $x_* = x^\dagger$, in which case theorem 2.8 guarantees convergence of $z_k \to x^\dagger$ (and then also $x_k \to x^\dagger$). Thus the above arguments apply to that case as well, which yields the assertion. □

We can now apply the above result to the TPG method (1.17) with constant stepsize $\alpha_k^\delta = \omega$ and λ_k^δ defined via (2.20). For this, we need the additional assumption

$$\sup_{x \in \mathcal{B}_{4\rho}(x_0)} \|F'(x)\| \leqslant \bar{\omega} < \infty. \tag{2.70}$$

Theorem 2.11. *Assume that (2.1) and (2.70) hold and that equation $F(x) = y$ has a solution x_* in $\mathcal{B}_\rho(x_0) = \mathcal{B}_\rho(x_{-1})$. Let $k_* = k_*(\delta, y^\delta)$ be chosen according to the discrepancy principle (2.13), (2.14). Assume that $\alpha_k^\delta = \omega \leqslant 1/\bar{\omega}^2$, where $\bar{\omega}$ satisfies (2.70) and that λ_k^δ is defined via (2.20), for some $\mu > 1$ and Ψ defined via (2.8). Then the iterates $z_{k_*}^\delta$ defined via (1.17) converge to a solution of $F(x) = y$, as $\delta \to 0$. If $\mathcal{N}(F'(x^\dagger)) \subset \mathcal{N}(F'(x))$ for all $x \in \mathcal{B}_{4\rho}(x^\dagger)$, then $z_{k_*}^\delta$ converges to $x^\dagger$ as $\delta \to 0$.*

Proof. Due to $\alpha_k^\delta = \omega \leqslant 1/\bar{\omega}^2$ and (2.70), there holds

$$\alpha_k^\delta \left\|s_k^\delta\right\|^2 \leqslant \left\|F(z_k^\delta) - y^\delta\right\|^2, \tag{2.71}$$

and hence, due to the discrepancy principle (2.13) and the definition of λ_k^δ via (2.20), we get that (2.15) is satisfied for all $0 \leqslant k < k_*$. Obviously, (2.4) and (2.34) hold, λ_k^δ depends continuously on δ for fixed k and, since $\lambda_k^0 = 0$, also (2.46) is trivially satisfied. Hence, theorem 2.10 is applicable, which immediately yields the desired results. □

3. Examples of TPG methods based on the steepest descent and the minimal error stepsize

In this section, we will introduce two TPG methods (1.17) based on the steepest descent and on the minimal error stepsize and show that, under some assumptions, they lead to convergent regularization methods. If we again denote

$$s_k^\delta := F'(z_k^\delta)^*(y^\delta - F(z_k^\delta)), \tag{3.1}$$

then the *steepest descent* stepsize α_k^{SD} is defined by

$$\alpha_k^{\mathrm{SD}} := \alpha_k^{\mathrm{SD}}(z_k^\delta) := \frac{\left\|s_k^\delta\right\|^2}{\left\|F'(z_k^\delta)s_k^\delta\right\|^2}, \tag{3.2}$$

and the *minimal error* stepsize α_k^{ME} is defined by

$$\alpha_k^{\mathrm{ME}} := \alpha_k^{\mathrm{ME}}(z_k^\delta) := \frac{\left\| y^\delta - F(z_k^\delta) \right\|^2}{\left\| s_k^\delta \right\|^2} . \tag{3.3}$$

The choice of the steepest descent stepsize α_k^{SD} is motivated by line-search procedures for optimization methods, where one tries to find an α_k^δ such that the functional

$$\frac{1}{2} \left\| F(z_k^\delta + \alpha_k^\delta s_k^\delta) - y^\delta \right\|^2 \tag{3.4}$$

is minimized. The stepsize α_k^{SD} minimizes the linearisation of this functional, i.e.

$$\alpha_k^{\mathrm{SD}} = \arg\min_{\alpha_k^\delta} \frac{1}{2} \left\| F(z_k^\delta) + \alpha_k^\delta F'(z_k^\delta) s_k^\delta - y^\delta \right\|^2 . \tag{3.5}$$

As for the minimal error stepsize α_k^{ME}, note that in the proof of proposition 2.2 we showed the following inequality:

$$\left\| x_{k+1}^\delta - x_* \right\|^2 \leqslant \left\| z_k^\delta - x_* \right\|^2 - \alpha_k^\delta \left(\left\| F(z_k^\delta) - y^\delta \right\|^2 - \alpha_k^\delta \left\| s_k^\delta \right\|^2 \right) . \tag{3.6}$$

Now, in order to ensure that $\left\| x_{k+1}^\delta - x_* \right\| \leqslant \left\| z_k^\delta - x_* \right\|$, the stepsize α_k^δ has to satisfy

$$\alpha_k^\delta \left\| s_k^\delta \right\|^2 \leqslant \left\| F(z_k^\delta) - y^\delta \right\|^2 , \tag{3.7}$$

and the choice of $\alpha_k^\delta = \alpha_k^{\mathrm{ME}}$ is the largest stepsize fulfilling that requirement.

In the following proposition we will show that α_k^{SD} and α_k^{ME} are well defined. The proof is almost completely similar to the one of [14, proposition 3.20].

Proposition 3.1. *Assume that* (2.1) *holds and that equation* $F(x) = y$ *has a solution* x_* *in* $\mathcal{B}_\rho(x_0)$. *Assume that* $x_k^\delta, x_{k-1}^\delta \in \mathcal{B}_\rho(x_*)$ *for an arbitrary* $k \in \mathbb{N} \cup \{0\}$ *and*

$$\left\| y^\delta - F(z_k^\delta) \right\| > 2 \frac{1+\eta}{1-2\eta} \delta \tag{3.8}$$

holds. Then $s_k^\delta \neq 0$ *and* $F'(z_k^\delta) s_k^\delta \neq 0$ *and consequently,* α_k^{SD} *and* α_k^{ME} *defined via* (3.2) *and* (3.3) *are well-defined.*

Proof. Since $x_k^\delta, x_{k-1}^\delta \in \mathcal{B}_\rho(x_*)$ it follows as in proposition 2.2 that $z_k \in \mathcal{B}_{4\rho}(x_0)$ and hence (2.1) is applicable. Assume now that $s_k^\delta = 0$. Then we have

$$\begin{aligned}
0 &= \left\langle s_k^\delta, z_k^\delta - x_* \right\rangle = \left\langle F'(z_k^\delta)^* (y^\delta - F(z_k^\delta)), z_k^\delta - x_* \right\rangle \\
&= \left\langle y^\delta - F(z_k^\delta), F'(z_k^\delta)(z_k^\delta - x_*) \right\rangle \\
&= \left\langle y^\delta - F(z_k^\delta), y^\delta - y + y - y^\delta + F(z_k^\delta) - F(z_k^\delta) + F'(z_k^\delta)(z_k^\delta - x_*) \right\rangle \\
&= \left\langle y^\delta - F(z_k^\delta), y^\delta - y \right\rangle - \left\| y^\delta - F(z_k^\delta) \right\|^2 \\
&\qquad - \left\langle y^\delta - F(z_k^\delta), F(z_k^\delta) - F(x_*) - F'(z_k^\delta)(z_k^\delta - x_*) \right\rangle .
\end{aligned} \tag{3.9}$$

Using (1.2) and (2.1), we get

$$\begin{aligned}
\left\| y^\delta - F(z_k^\delta) \right\|^2 &\leqslant \left\| y^\delta - F(z_k^\delta) \right\| \delta + \eta \left\| y^\delta - F(z_k^\delta) \right\| \left\| F(z_k^\delta) - F(x_*) \right\| \\
&\leqslant \left\| y^\delta - F(z_k^\delta) \right\| \delta + \eta \left\| y^\delta - F(z_k^\delta) \right\| \left(\left\| y^\delta - F(z_k^\delta) \right\| + \delta \right) \\
&= \left\| y^\delta - F(z_k^\delta) \right\| \left(\delta + \eta(\delta + \left\| y^\delta - F(z_k^\delta) \right\|) \right) ,
\end{aligned} \tag{3.10}$$

and therefore

$$\left\|y^\delta - F(z_k^\delta)\right\| \leqslant \frac{1+\eta}{1-\eta}\delta, \tag{3.11}$$

which is a contradiction to (3.8), hence $s_k^\delta \neq 0$.

Now assume that $F'(z_k^\delta)s_k^\delta = 0$. Then obviously $s_k^\delta \in \mathcal{N}(F'(z_k^\delta))$. By the definition of s_k^δ, we also have that $s_k^\delta \in \mathcal{R}(F'(z_k^\delta)^*) \subset \mathcal{N}(F'(z_k^\delta))^\perp$. Hence, we have $s_k^\delta = 0$, which is a contradiction to what we have shown above. Hence, $F'(z_k^\delta)s_k^\delta \neq 0$ and therefore α_k^{SD} and α_k^{ME} are well-defined. □

We now want to prove that all conditions on the stepsize α_k^δ used in the previous section also hold for α_k^{SD} and α_k^{ME}. We start by considering condition (2.34). Assuming (2.70) to hold, it then obviously follows that $\alpha_k^{\mathrm{SD}} \geqslant 1/\bar{\omega}^2$ and $\alpha_k^{\mathrm{ME}} \geqslant 1/\bar{\omega}^2$ and hence, condition (2.34) is satisfied. Now we state another helpful result due to [25].

Lemma 3.2. *For the stepsizes $\alpha_k^\delta = \alpha_k^{\mathrm{SD}}, \alpha_k^{\mathrm{ME}}$ defined via* (3.2) *and* (3.3)*, respectively, there holds*

$$\alpha_k^\delta \left\|s_k^\delta\right\|^2 \leqslant \left\|y^\delta - F(z_k^\delta)\right\|^2, \tag{3.12}$$

where equality holds for $\alpha_k^\delta = \alpha_k^{\mathrm{ME}}$ in the above inequality.

Proof. According to its definition, the statement is trivial for α_k^{ME}. For α_k^{SD}, it follows immediately from

$$\alpha_k^{\mathrm{SD}} \left\|s_k^\delta\right\|^2 = \frac{\left\langle F'(z_k^\delta)s_k^\delta, y^\delta - F(z_k^\delta)\right\rangle^2}{\left\|F'(z_k^\delta)s_k^\delta\right\|^2} \leqslant \left\|y^\delta - F(z_k^\delta)\right\|^2. \tag{3.13}$$

We now turn back to the very important condition (2.15). Due to lemma 3.2, if we use $\alpha_k^\delta = \alpha_k^{\mathrm{SD}}$ or $\alpha_k^\delta = \alpha_k^{\mathrm{ME}}$, then a sufficient condition for (2.15) to hold is given by

$$\lambda_k^\delta(\lambda_k^\delta + 1)\left\|x_k^\delta - x_{k-1}^\delta\right\|^2 \leqslant \frac{\Psi}{\mu}\alpha_k^\delta\left\|F(z_k^\delta) - y^\delta\right\|^2. \tag{3.14}$$

As we previously noted in section 2, the choice $\lambda_k^\delta = 0$ satisfies this inequality, which, however, corresponds to the classical steepest descent or minimal error method, respectively. Another possibility which, using (2.70), can be derived analogously to (2.20), is given by

$$\lambda_k^\delta = \min\left\{-\frac{1}{2} + \sqrt{\frac{1}{4} + \frac{\Psi(\tau\delta)^2}{\mu\,\bar{\omega}^2\left\|x_k^\delta - x_{k-1}^\delta\right\|^2}}, 1\right\}. \tag{3.15}$$

Note that this is the same as (2.20), given that the optimal stepsize $\omega = 1/\bar{\omega}^2$ is being used. For λ_k^δ as in (3.15), we can deduce the following.

Theorem 3.3. *Assume that* (2.1) *and* (2.70) *hold and that equation $F(x) = y$ has a solution x_* in $\mathcal{B}_\rho(x_0) = \mathcal{B}_\rho(x_{-1})$. Let $k_* = k_*(\delta, y^\delta)$ be chosen according to the discrepancy* *principle* (2.13), (2.14). *Assume that either* $\alpha_k^\delta = \alpha_k^{\mathrm{SD}}$ *or* $\alpha_k^\delta = \alpha_k^{\mathrm{ME}}$, *defined by* (3.2) *or* (3.3), *respectively. Furthermore, let* λ_k^δ *be defined via* (3.15), *for some* $\mu > 1$, Ψ *defined via* (2.8) *and* $\bar{\omega}$ *satisfying* (2.70). *Then the iterates* $z_{k_*}^\delta$ *defined via* (1.17) *converge to a solution of* $F(x) = y$, *as* $\delta \to 0$. *If* $\mathcal{N}(F'(x^\dagger)) \subset \mathcal{N}(F'(x))$ *for all* $x \in \mathcal{B}_{4\rho}(x^\dagger)$, *then* $z_{k_*}^\delta$ *converges to* $x^\dagger$ *as* $\delta \to 0$.

Proof. From lemma 3.2, we get that

$$\alpha_k^\delta \left\| s_k^\delta \right\|^2 \leqslant \left\| F(z_k^\delta) - y^\delta \right\|^2 . \tag{3.16}$$

Together with $\alpha_k^{\mathrm{SD}}, \alpha_k^{\mathrm{ME}} \geqslant 1/\bar{\omega}^2$, the statements of the theorem now follow from theorem 2.10, analogously as in the proof of theorem 2.11. □

As for λ_k^δ defined via (2.20), for λ_k^δ defined via (3.15) there also holds $\lambda_k^\delta = 0$ for $\delta = 0$. Since this corresponds to classical Landweber iteration, the steepest descent or minimal error method, the acceleration effect due to those choices of λ_k^δ will decrease for $\delta \to 0$. Since for small values of δ acceleration is needed most, other choices of λ_k^δ also have to be considered.

The crucial conditions which a pair $(\lambda_k^\delta, \alpha_k^\delta)$ has to satisfy in order for theorem 2.10 to be applicable are the conditions (2.15) and (2.46). We have already seen that $\lambda_k^\delta = 0$ and λ_k^δ defined via either (2.20) or (3.15), and hence, all sequences in between those two, satisfy the coupling condition (2.15). Given a stepsize α_k^δ, one could think of choosing $\lambda_k^\delta \leqslant 1$ as large as possible such that the coupling condition (2.15) is satisfied. However, one also has to guarantee that condition (2.46) is satisfied as well.

One possibility is to choose λ_k^δ as a subsequence of a summable sequence like $(cq^k)_{k\in\mathbb{N}}$, $0 \leqslant q < 1$, in such a way that (2.15) is satisfied, which, together with the boundedness of $\left\| x_k^\delta - x_{k-1}^\delta \right\|$, guarantees (2.46). Unfortunately, the resulting sequence λ_k^δ tends to 0 as $k \to \infty$, which in turn only leads to a negligible acceleration effect. However, notice that for condition (2.46) to be satisfied, it suffices that the sequence $\lambda_k^0 \left\| x_k - x_{k-1} \right\|$ is summable. Hence, we propose the following strategy.

Given a stepsize α_k^δ, define the combination parameters λ_k^δ via

$$\lambda_k^\delta = \begin{cases} 0, & k = 0, \\ \min\left\{ \frac{q_k^\delta}{\left\| x_k^\delta - x_{k-1}^\delta \right\|}, 1 \right\}, & k \geqslant 1, \end{cases} \tag{3.17}$$

where $(q_k^\delta)_{k\in\mathbb{N}}$ is a decreasing sequence depending continuously on δ for fixed k, satisfying

$$\sum_{k=0}^{\infty} q_k^\delta < \infty, \tag{3.18}$$

and chosen such that condition (2.15) holds. If the sequence $(q_k^\delta)_{k\in\mathbb{N}}$ can be chosen in such a way that it converges to zero fast enough to satisfy (3.18) but slower than $\left\| x_k^\delta - x_{k-1}^\delta \right\|$, the resulting sequence λ_k^δ will stay away from zero and possibly even tend towards one as $k \to \infty$.

Finding a sequence $(q_k^\delta)_{k\in\mathbb{N}}$ satisfying all the required properties such that the resulting TPG method indeed gives rise to a convergent regularization method and how to compute a viable sequence λ_k^δ in practise will be the topics of the remainder of this section. First, we will consider the problem of finding a suitable sequence $(q_k^\delta)_{k\in\mathbb{N}}$, or alternatively, λ_k^δ, via what in the following we will call the *backtracking search* (BTS) *algorithm*, given by:

Algorithm 3.1. BTS algorithm for λ_k^δ, $k > 1$.

- **Given:** x_k^δ, x_{k-1}^δ, Ψ, μ, y^δ, F, $q : \mathbb{R}_0^+ \to \mathbb{R}_0^+$, $m_{k-1}^\delta \in \mathbb{R}$.
- Calculate $\left\| x_k^\delta - x_{k-1}^\delta \right\|$ and define

$$\beta_k^\delta(m) := \min\left\{ \frac{q(m)}{\left\| x_k^\delta - x_{k-1}^\delta \right\|}, 1 \right\} . \tag{3.19}$$

- Define the functions

$$\tilde{\lambda}_k^\delta(m) := \beta_k^\delta(m_{k-1}^\delta + 1 + m) ,$$
$$\tilde{z}_k^\delta(m) := x_k^\delta + \tilde{\lambda}_k^\delta(m)(x_k^\delta - x_{k-1}^\delta) ,$$
$$\tilde{\alpha}_k^\delta(m) := \alpha_k^\delta(\tilde{z}_k^\delta(m)) . \tag{3.20}$$

- Calculate

$$\tilde{m}_k^\delta = \inf\left\{ m \geqslant 0 \,\middle|\, \tilde{\lambda}_k^\delta(m)(\tilde{\lambda}_k^\delta(m)+1) \left\| x_k^\delta - x_{k-1}^\delta \right\|^2 \leqslant \tfrac{\Psi}{\mu} \tilde{\alpha}_k^\delta(m) \left\| y^\delta - F(\tilde{z}_k^\delta(m)) \right\|^2 \right\} . \tag{3.21}$$

- Define $\lambda_k^\delta := \tilde{\lambda}_k^\delta(\tilde{m}_k^\delta)$, $z_k^\delta := \tilde{z}_k^\delta(\tilde{m}_k^\delta)$ and $m_k^\delta := m_{k-1}^\delta + 1 + \tilde{m}_k^\delta$.
- **Output:** λ_k^δ, z_k^δ, m_k^δ.

In order to carry out the above algorithm, a function $q : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ needs to be specified. In order to prove convergence of our iteration method with λ_k^δ chosen via algorithm 3.1, we will have to make the following assumptions on this function:

$$q(m_1) \leqslant q(m_2) \quad \forall\, m_1 > m_2, \qquad \sum_{k=0}^{\infty} q(k) < \infty. \tag{3.22}$$

Concerning the calculation of $\tilde{m}_k^\delta$, note first that it is possible that $\tilde{\alpha}_k^\delta(m)$ is not well-defined for certain values of m. However, by proposition 3.1 this can only happen if $\tilde{z}_k^\delta(m)$ is such that (3.8) holds, i.e. that the stopping criterion (2.13) is satisfied, and we will therefore consider the inequality in (3.21) to be satisfied for those m. Furthermore, if there is no $m \geqslant 0$ such that the inequality

$$\tilde{\lambda}_k^\delta(m)(\tilde{\lambda}_k^\delta(m)+1) \left\| x_k^\delta - x_{k-1}^\delta \right\|^2 \leqslant \frac{\Psi}{\mu} \tilde{\alpha}_k^\delta(m) \left\| y^\delta - F(\tilde{z}_k^\delta(m)) \right\|^2 \tag{3.23}$$

is satisfied, then $\tilde{m}_k^\delta = \inf \emptyset = \infty$ and hence $\tilde{\lambda}_k^\delta(\tilde{m}_k^\delta)$ and $\tilde{z}_k^\delta(\tilde{m}_k^\delta)$ have to be understood in the limit sense, i.e.

$$\tilde{\lambda}_k^\delta(\infty) := \lim_{m\to\infty} \tilde{\lambda}_k^\delta(m) = 0, \qquad \tilde{z}_k^\delta(\infty) := \lim_{m\to\infty} \tilde{z}_k^\delta(m) = x_k^\delta. \tag{3.24}$$

However, since by (3.19) and (3.22) there holds $\tilde{\lambda}_k^\delta(m) \to 0$ as $m \to 0$ and since α_k^δ is bounded away from 0 in this case, $\tilde{m}_k^\delta = \infty$ can only happen if $\left\| y^\delta - F(\tilde{z}_k^\delta(m)) \right\| \to 0$ as $m \to \infty$. By the continuity of the involved quantities, this in turn implies $\left\| y^\delta - F(\tilde{z}_k^\delta(\infty)) \right\| = 0$ and hence, due to the discrepancy principle, the TPG method will be terminated with $z_k^\delta = \tilde{z}_k^\delta(\infty)$ after the current iteration.

Combining the above considerations, for TPG methods (1.17) combined with the BTS algorithm 3.1 for determining a suitable sequence λ_k^δ we can now prove the following convergence result.

Theorem 3.4. *Assume that* (2.1) *and* (2.70) *hold and that equation* $F(x) = y$ *has a solution* x_* *in* $\mathcal{B}_\rho(x_0) = \mathcal{B}_\rho(x_{-1})$. *Let* x_k^δ, z_k^δ *be defined via* (1.17) *with* α_k^δ *being given by either* (3.2) *or* (3.3). *Let* λ_k^δ *be defined via algorithm* 3.2 *with* $\lambda_0^\delta = 0$, $m_0^\delta = 0$, $\mu > 1$, Ψ *as in* (2.8) *and* $q : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ *satisfying* (3.22). *Let* $k_* = k_*(\delta, y^\delta)$ *be chosen according to the discrepancy principle* (2.13), (2.14). *Then the following statements hold:*

1. *if* $y = y^\delta$, *i.e. if* $\delta = 0$, *and if* $k_* = k_*(0, y) = \infty$ *then the iterates* z_k *and* x_k *converge to a solution of* $F(x) = y$ *as* $k \to \infty$. *If* $\mathcal{N}(F'(x^\dagger)) \subset \mathcal{N}(F'(x))$ *for all* $x \in \mathcal{B}_{4\rho}(x^\dagger)$, *then* z_k *and* x_k *converge to* $x^\dagger$ *as* $k \to \infty$.
2. *For all* $(-1) \leqslant k < k_*$ *there holds* $\left\|x_{k+1}^\delta - x_*\right\| \leqslant \left\|x_k^\delta - x_*\right\|$. *Furthermore, if, for fixed* k, $\tilde{m}_k^\delta$ *defined via* (3.21) *depends continuously on the data as* $\delta \to 0$ *then* $z_{k_*}^\delta$ *converges to a solution of* $F(x) = y$ *as* $\delta \to 0$. *If additionally* $\mathcal{N}(F'(x^\dagger)) \subset \mathcal{N}(F'(x))$ *for all* $x \in \mathcal{B}_{4\rho}(x^\dagger)$, *then* $z_{k_*}^\delta$ *converges to* $x^\dagger$ *as* $\delta \to 0$.

Proof. From algorithm 3.1 it is obvious that $m_k^\delta \geqslant m_{k-1}^\delta + 1$ and therefore $m_k^\delta \geqslant k$. Using this together with (3.22), we get that

$$\begin{aligned} \sum_{k=0}^{\infty} \lambda_k^0 \left\|x_k - x_{k-1}\right\| &\leqslant \sum_{k=0}^{\infty} \beta_k^0(m_k^0) \left\|x_k - x_{k-1}\right\| = \sum_{k=0}^{\infty} \min\left\{q(m_k^0), \left\|x_k - x_{k-1}\right\|\right\} \\ &\leqslant \sum_{k=0}^{\infty} q(m_k^0) \leqslant \sum_{k=0}^{\infty} q(k) < \infty, \end{aligned} \tag{3.25}$$

from which it follows that (2.46) holds. Furthermore, condition (3.14) follows directly from the definition of $\lambda_k^\delta = \tilde{\lambda}_k^\delta(\tilde{m}_k^\delta)$ and due to (3.19), also $0 \leqslant \lambda_k^\delta \leqslant 1$ holds. Together with the observations made above, the first part of this theorem follows immediately from theorem 2.8 and corollary 2.9, as does the monotonicity result in the second part of the theorem. Furthermore, if $\tilde{m}_k^\delta$ depends continuously on the data, i.e. if, for fixed k, $\tilde{m}_k^\delta \to \tilde{m}_k^0$ as $\delta \to 0$, then by the continuity of the involved quantities, also the sequence λ_k^δ defined via algorithm 3.1 depends continuously on δ for $\delta \to 0$ and fixed k. Using this, the remaining statements of the theorem now follow immediately from theorem 2.10. □

Concerning the convergence analysis above, note that we require that $\tilde{m}_k^\delta$ depends continuously on δ as $\delta \to 0$. Comparing this with the definition (3.21) of $\tilde{m}_k^\delta$, we see that it is equivalent to requiring that the first point of intersection of the two functions

$$f^\delta(m) := \tilde{\lambda}_k^\delta(m)(\tilde{\lambda}_k^\delta(m) + 1) \left\|x_k^\delta - x_{k-1}^\delta\right\|^2 \quad \text{and} \quad g^\delta(m) := \frac{\Psi}{\mu}\tilde{\alpha}_k^\delta(m) \left\|y^\delta - F(\tilde{z}_k^\delta(m))\right\|^2$$

depends continuously on δ as $\delta \to 0$. Although this might not always necessarily be true due to pathological cases, it is reasonable to expect this to be true in practise.

The BTS algorithm 3.1 has one disadvantage, namely the fact that one has to calculate an infimum for determining $\tilde{m}_k^\delta$. While this might be possible analytically for very specific problems, in general one cannot hope to be able to resolve the infimum explicitly. In order to avoid having to approximate this infimum numerically via some potentially very costly numerical routine, we introduce a numerically feasible version of the BTS algorithm, which we will call *discrete backtracking search* (DBTS) *algorithm*. It is based on the same ideas as the BTS algorithm and takes the following form:

Algorithm 3.2. DBTS algorithm for λ_k^δ, $k > 1$.

- **Given:** x_k^δ, x_{k-1}^δ, τ, Ψ, μ, y^δ, F, $q : \mathbb{R}_0^+ \to \mathbb{R}_0^+$, $i_{k-1} \in \mathbb{N}$, $j_{\max} \in \mathbb{N}$.
- Calculate $\left\|x_k^\delta - x_{k-1}^\delta\right\|$ and define

$$\beta_k(i) = \min\left\{\frac{q(i)}{\left\|x_k^\delta - x_{k-1}^\delta\right\|}, 1\right\}. \tag{3.26}$$

- **For:** $j = 1 \ldots, j_{\max}$,

  Set $\lambda_k^\delta = \beta_k(i_{k-1} + j)$.
  Calculate $z_k^\delta = x_k^\delta + \lambda_k^\delta(x_k^\delta - x_{k-1}^\delta)$ and $\alpha_k^\delta = \alpha_k^\delta(z_k^\delta)$.
  If: $(\left\|y^\delta - F(z_k^\delta)\right\| \leqslant \tau\delta)$,
    $i_k = i_{k-1} + j$,
    break.
  Elseif: $\left(\lambda_k^\delta(\lambda_k^\delta + 1)\left\|x_k^\delta - x_{k-1}^\delta\right\|^2 \leqslant \frac{\Psi}{\mu}\alpha_k^\delta\left\|y^\delta - F(z_k^\delta)\right\|^2\right)$,
  $i_k = i_{k-1} + j$,
    break.
  Else: $\lambda_k^\delta = 0$, $i_k = i_{k-1} + j_{\max}$.

End For

- **Output:** λ_k^δ, i_k.

The above algorithm is easy to implement and does not require the computation of an infimum. Furthermore, similarly to above we can show a convergence result.

Theorem 3.5. *Assume that (2.1) and (2.70) hold and that equation $F(x) = y$ has a solution x_* in $\mathcal{B}_\rho(x_0) = \mathcal{B}_\rho(x_{-1})$. Let x_k^δ, z_k^δ be defined via (1.17) with α_k^δ being given by either (3.2) or (3.3). Let λ_k^δ be defined via algorithm 3.1 with $\lambda_0^\delta = 0$, $j_{\max} \in \mathbb{N}$, $\mu > 1$, τ as in (2.14), Ψ as in (2.8) and $q : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ satisfying (3.22). Let $k_* = k_*(\delta, y^\delta)$ be chosen according to the discrepancy principle (2.13), (2.14). Then there holds:*

1. *if $y = y^\delta$, i.e. if $\delta = 0$, and if $k_* = k_*(0, y) = \infty$ then the iterates z_k and x_k converge to a solution of $F(x) = y$ as $k \to \infty$. If $\mathcal{N}(F'(x^\dagger)) \subset \mathcal{N}(F'(x))$ for all $x \in \mathcal{B}_{4\rho}(x^\dagger)$, then z_k and x_k converge to $x^\dagger$ as $k \to \infty$.*
2. *For all $(-1) \leqslant k < k_*$ there holds $\left\|x_{k+1}^\delta - x_*\right\| \leqslant \left\|x_k^\delta - x_*\right\|$. Furthermore, if $k_*(0, y) = \infty$ and if for all $k \in \mathbb{N}$ there holds*

$$\lambda_k^0(\lambda_k^0 + 1)\left\|x_k - x_{k-1}\right\|^2 < \frac{\Psi}{\mu}\alpha_k^0\left\|y - F(z_k)\right\|^2, \tag{3.27}$$

   then $z_{k_*}^\delta$ converges to a solution of $F(x) = y$ as $\delta \to 0$. If additionally $\mathcal{N}(F'(x^\dagger)) \subset \mathcal{N}(F'(x))$ for all $x \in \mathcal{B}_{4\rho}(x^\dagger)$, then $z_{k_*}^\delta$ converges to $x^\dagger$ as $\delta \to 0$.

Proof. The proof of this theorem is analogous to the proof of theorem 3.4. Note that due to checking whether $\left\|y^\delta - F(z_k^\delta)\right\| \leqslant \tau\delta$, the stepsize α_k^δ is guaranteed to be well defined during the search procedure and the iteration. Furthermore, the assumption that $k_*(0, y) = \infty$ together with (3.27) and the continuity of the involved quantities implies that for fixed k, $\lambda_k^\delta \to \lambda_k^0$ as $\delta \to 0$. $\square$

Note that the analysis carried out above in theorems 3.4 and 3.5 also applies to constant stepsizes $\alpha_k^\delta = \omega$, as long as $\omega \leqslant 1/\bar{\omega}^2$ with $\bar{\omega}$ satisfying (2.70), since for that choice, as we have already seen in the proof of theorem 2.11, the results of lemma 3.2 hold as well. Furthermore, in this case, the If branch in the DBTS algorithm which checks whether $\|y^\delta - F(z_k^\delta)\| \leqslant \tau\delta$ can be dropped, since the stepsize is now always well-defined. Consequently, also the requirement that $k_*(0, y) = \infty$ in the second part of theorem 3.5 can then be removed. Hence, using a TPG method with a constant stepsize combined with the BTS algorithm for λ_k^δ gives rise to a convergent regularization method as well.

Note that in order to apply either of the backtracking search algorithms presented above one needs to have an estimate of the same parameters as for ordinary nonlinear Landweber iteration, that is, of δ and η. Whereas in ordinary Landweber iteration η only plays a role in choosing τ, here it also enters into the BTS and DBTS algorithms via Ψ. For linear problems, $\eta = 0$ can be chosen and therefore

$$\Psi = 1 - 2\tau^{-1}, \quad \text{with} \quad \tau > 2. \tag{3.28}$$

If we take for example $\tau = 4$, then we get $\Psi = 1/2$. Note that one would want to have τ as small and Ψ as big as possible. However, since by the above equation τ and Ψ are directly proportional, one has to settle for a compromise when choosing τ. Note also that usually the exact value of η is not known. In this case, a value for η close to 0.5 is chosen in numerical algorithms requiring η explicitly.

4. Numerical examples

In this section, we numerically demonstrate the acceleration effect of our proposed TPG methods (1.17) compared to their non-accelerated counterparts. We do this by looking at a nonlinear Hammerstein operator and at the 2D inverse problem of SPECT.

4.1. Numerical example—nonlinear Hammerstein operator

As a first example, we consider the nonlinear Hammerstein integral operator

$$F : H^1[0,1] \to L^2[0,1], \quad F(x)(s) := \int_0^s (x(t))^3 \, \mathrm{d}t, \tag{4.1}$$

which is often used in the literature (see for example [9, 17–19]) to illustrate convergence conditions, demonstrate convergence rates and show the effects of different stepsizes and acceleration techniques. Importantly, the operator F is Fréchet differentiable and furthermore, if $x \geqslant \kappa > 0$ for all $x \in \mathcal{B}_{4\rho}(x_0)$ then one can show that there exists a family of bounded linear operators $R_x(\tilde{x}) : \mathcal{Y} \to \mathcal{Y}$ and a constant $c > 0$ such that

$$F'(x) = R_x(\tilde{x})F'(\tilde{x}), \qquad \|R_x(\tilde{x}) - I\| \leqslant c\,\|x - \tilde{x}\|\,, \tag{4.2}$$

for all $x, \tilde{x} \in \mathcal{B}_{4\rho}(x_0) \subset \mathcal{D}(F)$, which in particular implies that

$$\|F(x) - F(\tilde{x}) - F'(\tilde{x})(x - \tilde{x})\| \leqslant \frac{c}{2 - c\,\|x - \tilde{x}\|}\,\|x - \tilde{x}\|\,\|F(x) - F(\tilde{x})\|\,. \tag{4.3}$$

Hence, if $x^\dagger \in \mathcal{B}_\rho(x_0)$ satisfies $x^\dagger \geqslant \bar{\kappa} > 0$ and if $\rho > 0$ is small enough such that both $x \geqslant \kappa > 0$ for all $x \in \mathcal{B}_{4\rho}(x_0)$ and $6c\rho < 1$ are satisfied, then the nonlinearity condition (2.1) holds with

$$\eta = \frac{2c\rho}{1-2c\rho} < \frac{1}{2}. \tag{4.4}$$

Hence, since for this problem the operators $R_x(\tilde{x})$ can be given explicitly by (see [9])

$$R_x(\tilde{x})^* w = -\left(\frac{\phi'(x)}{\phi'(\tilde{x})}\int_{\bullet}^{1} w(t)\,\mathrm{d}t\right)', \tag{4.5}$$

it is possible to determine an η from (4.4) by deriving an estimate of the constant c in (4.2). Since explicit estimates of this constant are usually not sharp enough, one often tries to numerically compute an estimate for c. However, since we do not require c but only η for our tests, we will numerically estimate η directly from (2.1).

For our tests we use the same setup as in [19], i.e. we assume that $y = F(x^\dagger)$ with

$$x^\dagger(t) := 1 + 10^{-2}(7 - 3t^2 + 2t^3), \tag{4.6}$$

and that $x_0(t) = 1$. Hence, we have that

$$x^\dagger - x_0 \in \mathcal{R}(F'(x^\dagger)^*) \quad \text{and} \quad \rho = \left\|x^\dagger - x_0\right\| = \frac{1}{100}\sqrt{\frac{305}{7}} \approx 0.066. \tag{4.7}$$

Numerical calculations show that the constant c in (4.2) is given by $c \approx 3$, which, by (4.4) would imply that $\eta \approx 0.656 > \frac{1}{2}$. However, numerically estimating η directly via (2.1) shows that η is actually much smaller, i.e. $\eta \approx 0.4$. Moreover, when using classical Landweber iteration, with or without the steepest descent or the minimal error stepsize, condition (2.1) only has to hold on $\mathcal{B}_{2\rho}(x_0)$ (see [14]). Estimating η on this set gives $\eta \approx 0.2$, the choice of which leads to strongly improved results also for our TPG methods. Hence, we will use $\eta = 0.2$ in all of the numerical tests below.

In order to discretize the problem, we subdivide the interval $[0, 1]$ into $n = 128$ equally spaced subintervals and replace the operators F, $F'(x)$ and $F'(x)^*$ by finite dimensional approximations defined in the same way as in [17, 19]. The data was created on a finer grid and a random relative data error of 0.001% was added to get y^δ.

We now want to compare the TPG methods based on a constant stepsize ω, the steepest descent stepsize α_k^{SD} and the minimal error stepsize α_k^{ME}, which we introduced in the previous section, with their classical, non-accelerated counterparts. For choosing λ_k^δ, we will use the Nesterov combination parameter (compare with (1.16)),

$$\lambda_k^N := \frac{k-1}{k+\alpha-1}, \tag{4.8}$$

where we will only consider the standard choice $\alpha = 3$, the sequence of λ_k^δ defined via the DBTS algorithm 3.2, which we will denote by λ_k^B, as well as the sequences given explicitly by (2.20) and (3.15), which are equivalent, since we will use $\omega = 1/\bar{\omega}^2$, and which we will denote by λ_k^E.

For using the DBTS algorithm, but also for choosing a suitable τ in the discrepancy principle, the approximation for η described above was used. From this, Ψ was calculated via (2.8) and τ was chosen via

$$\tau = 2\tilde{\tau}\frac{1+\eta}{1-2\eta}, \tag{4.9}$$

where $\tilde{\tau} = 1.01$, which ensures that condition (2.14) is satisfied. In the backtracking algorithm for λ_k^B, we use $j_{\max} = 5$ and $\mu = 2$. For the function $q : \mathbb{R}_0^+ \to \mathbb{R}_0^+$, we use $q(m) = 1/m^{1.1}$, which obviously satisfies the necessary condition (3.22). When using a constant stepsize, we use the scaling parameter $\omega = 0.3175$, which is chosen via numerically estimating the constant $\bar{\omega}$ in (2.70) and then taking $\omega = 1/\bar{\omega}^2$.

A summary of the results can be found in table 1. For both the constant and the steepest descent stepsize all three non-zero combination parameters λ_k^δ lead to a considerable decrease in the required number of iterations and computation time to meet the stopping rule. The choices $\lambda_k^\delta = \lambda_k^E$ and $\lambda_k^\delta = \lambda_k^B$ seem to perform equally well, with the explicit choice $\lambda_k^\delta = \lambda_k^E$ requiring slightly less time and iterations in both cases. Furthermore, using the combination parameter $\lambda_k^\delta = \lambda_k^N$ requires the least amount of time and iterations, the necessary time being more than halved in the case of the steepest descent stepsize. For the minimal error stepsize, the choice $\lambda_k^\delta = \lambda_k^N$ is again the best of all three non-zero combination parameters $\lambda_k^\delta = 0$. However, using $\lambda_k^\delta = 0$, i.e. the pure minimal error method without acceleration, only seven iterations are required, making it the best reconstruction method for this example. This fact was already observed in [19], where regardless of the discretization and the noise level, a constant number of iterations was required to meet the stopping rule. No explanation for this could be given in [19] for this pathological case and here we only state that in the numerical example treated in the next section, the choice $\lambda_k^\delta = \lambda_k^N$ requires significantly less iterations than the choice $\lambda_k^\delta = 0$ also for the minimal error stepsize.

4.2. Numerical example—SPECT

In the medical imaging technique of SPECT, one aims at reconstructing a radioactive distribution f, termed *activity function*, from radiation measurements outside the body, denoted by y. The usual modelling approach connects f and y via the *attenuated Radon transform* (ATRT), see for example [15], which is given by

$$y = A(f,\mu)(s,\omega) := \int_{\mathbb{R}} f(s\omega^\perp + t\omega)\exp\left(-\int_t^\infty \mu(s\omega^\perp + r\omega)\,\mathrm{d}r\right)\mathrm{d}t, \tag{4.10}$$

where $s \in \mathbb{R}, \omega \in S^1$. The function μ is called an *attenuation map* and is related to the density of different tissues. If μ is known, then reconstructing f from y is a linear problem. However, unless an additional CT (computerized tomography) scan is performed, which is not preferable due to the increased cost of the medical examination, μ is unknown as well. Hence, we face the nonlinear inverse problem of reconstructing the pair (f,μ) from y, or rather, from a noisy version y^δ of y.

This inverse problem and its numerical treatment, under various additional conditions like sparsity, has already been extensively studied (see for example [4, 5, 22, 23] and the references therein). Considering the definition space of the ATRT operator, it was shown in [4], that if

$$\mathcal{D}(A) := H_0^{s_1}(\Omega) \times H_0^{s_2}(\Omega), \tag{4.11}$$

where $H_0^s(\Omega)$ is the classical Sobolev space of order s over the bounded domain Ω with zero boundary conditions, then, assuming that the chosen s_1 and s_2 are large enough, the operator A is twice the continuous Fréchet differentiable with a Lipschitz continuous first derivative. Since one expects some discontinuities in (f,μ), one wants to choose s_1 and s_2 as small as possible. In [4] it was shown that it is possible to use $s_1 > 4/9$ and $s_2 = 1/3$, a choice which also allows a certain amount of non-smoothness of (f,μ).

For our numerical simulations, we used the so-called MCAT-phantom [10], which is depicted in figure 1. As one can see, the simulated activity function f_* is concentrated in the heart and the attenuation function μ_* models a cut through the thorax. Both functions are

Table 1. Comparison of different stepsizes α_k^δ and combination parameters λ_k^δ: number of iterations k_* and total amount of time necessary to satisfy the discrepancy principle. A relative data error of 0.001% was used.

Stepsize	$\lambda_k^\delta = 0$	$\lambda_k^\delta = \lambda_k^E$	$\lambda_k^\delta = \lambda_k^B$	$\lambda_k^\delta = \lambda_k^N$	k_*	Time (s)
Steepest descent	x				125	79
Steepest descent		x			35	22
Steepest descent			x		41	26
Steepest descent				x	14	9
Minimal error	x				7	4
Minimal error		x			183	116
Minimal error			x		192	135
Minimal error				x	78	45
Constant, $\omega = 0.3175$	x				260	178
Constant, $\omega = 0.3175$		x			42	29
Constant, $\omega = 0.3175$			x		48	33
Constant, $\omega = 0.3175$				x	32	22

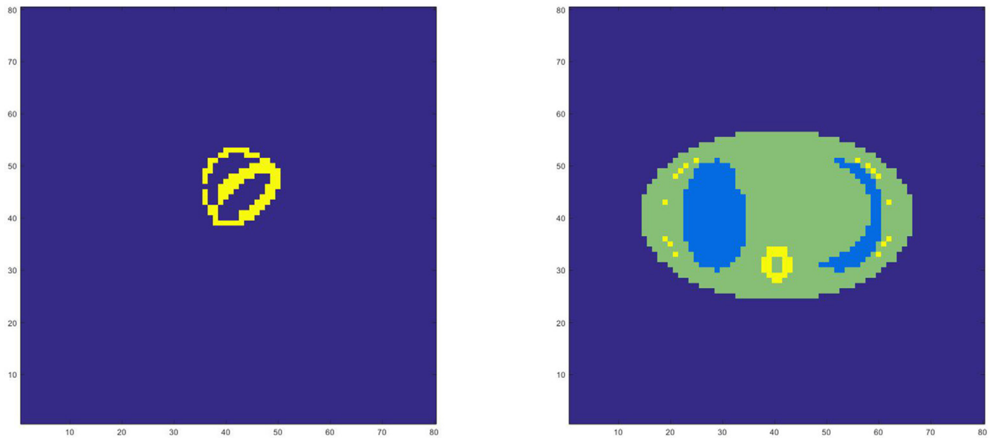

Figure 1. Activity function f_* (left) and attenuation function μ_* (right).

given as 80×80 pixel images. The Radon transform, its Fréchet derivative and the adjoint thereof were discretized to work on those pixel images, using 79 angles ω, equally distributed over 360 degrees, and 80 samples for s.

The data y was calculated via $y = A(f_*, \mu_*)$, i.e. by applying the discretized version of the attenuated Radon transform to the pair (f_*, μ_*). The resulting sinogram is depicted in figure 2, once for the already shown attenuation function μ_* and once for $\mu_* = 0$. Afterwards, random data error was added in order to arrive at y^δ.

As in the previous section, we now want to compare the TPG methods based on a constant stepsize ω, the steepest descent stepsize α_k^{SD} and the minimal error stepsize α_k^{ME} with their classical, non-accelerated counterparts. Again we use the notation λ_k^E, λ_k^B and λ_k^N to distinguish between the different combination parameters λ_k^δ.

Concerning the nonlinearity constant η, it is not clear weather a condition like (2.1) holds for SPECT. Unfortunately, this is the case for almost all nonlinear inverse problems of practical

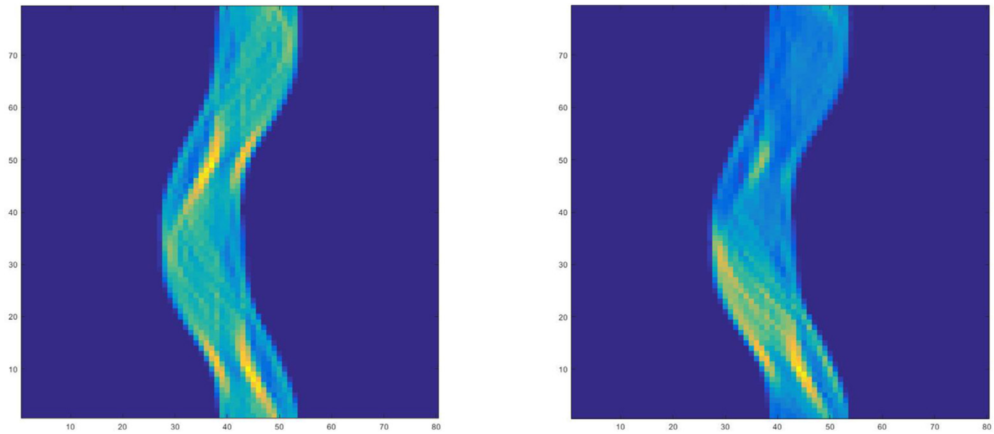

Figure 2. The generated data $y = A(f_*, 0)$ (left) and $y = A(f_*, \mu_*)$ (right).

importance. However, a value for η is both in the DBTS algorithm and for calculating Ψ and τ. Hence, we used the conservative estimate of $\eta = 0.4$ for obtaining the presented results. From this, Ψ was calculated via (2.8) and τ was chosen via

$$\tau = 2\tilde{\tau}\frac{1+\eta}{1-2\eta}, \tag{4.12}$$

where this time $\tilde{\tau} = 4$ was chosen. The resulting $\tau = 56$ might seem rather large but numerical tests show that decreasing τ for example to the canonical choice $\tau = 2$ leads to numerical instabilities which make it impossible for any of the methods to decrease the residual to the level of $\tau\delta$. Hence, the choice of τ as stated above seems to be at least of optimal order. Furthermore, as noted in the last paragraph of section 3, τ should not be chosen too small since otherwise Ψ would become undesirably small. Concerning the remaining parameters, they were all chosen as in the previous section, with the obvious exception of ω, for which the value $\omega = 4.7 \cdot 10^{-4}$ was found by numerical calculations.

We now compare the effects of combining different choices of λ_k^δ with different stepsizes α_k^δ. For this test, the results of which are presented in table 2, we used a relative data error of 0.25%[5]. Note first that independently of the chosen stepsize α_k^δ, using $\lambda_k^\delta = \lambda_k^N$ leads to the smallest number of iterations necessary before meeting the stopping rule, with only about one tenth of iterations and computation time required! For $\lambda_k^\delta = \lambda_k^B$ defined via the DBTS algorithm, we can see that for the constant stepsize $\omega = 10^{-5}$ and the steepest descent stepsize α_k^{SD}, although requiring more iterations and computation time, the overall effort is still significantly lower than when not using any acceleration. The bad behaviour of the combination of λ_k^B with the minimal error stepsize α_k^{ME} can best be explained by the fact that using the minimal error stepsize, the residuals are not decreasing monotonously and hence, the DBTS algorithm has difficulties finding a suitable parameter λ_k^B. As for the choice $\lambda_k^\delta = \lambda_k^E$, one can see that in combination with the steepest descent stepsize α_k^{SD}, about three times as many iterations are required than when using $\lambda_k^\delta = \lambda_k^N$. However, still much less iterations are required than when using no acceleration at all. A similar phenomenon can also be observed for the constant stepsize ω, where the choice $\lambda_k^\delta = \lambda_k^E$ can even compete with the choice

[5] This is a very optimistic estimate for SPECT, since in practice one would expect the relative data error to be upwards of 5%. However, for such a high amount of noise, only a couple of iterations are required to satisfy the stopping criterion (2.13) even for Landweber iteration and hence, no acceleration effect would be observable.

Table 2. Comparison of different stepsizes α_k^δ and combination parameters λ_k^δ: number of iterations k_* and total amount of time necessary to satisfy the discrepancy principle. A relative data error of 0.25% was used.

Stepsize	$\lambda_k^\delta = 0$	$\lambda_k^\delta = \lambda_k^E$	$\lambda_k^\delta = \lambda_k^B$	$\lambda_k^\delta = \lambda_k^N$	k_*	Time (s)
Steepest descent	x				3433	489
Steepest descent		x			631	90
Steepest descent			x		345	77
Steepest descent				x	205	30
Minimal error	x				2021	185
Minimal error		x			6665	603
Minimal error			x		6253	600
Minimal error				x	288	28
Constant, $\omega = 4.7 \cdot 10^{-4}$	x				2019	186
Constant, $\omega = 4.7 \cdot 10^{-4}$		x			474	46
Constant, $\omega = 4.7 \cdot 10^{-4}$			x		467	57
Constant, $\omega = 4.7 \cdot 10^{-4}$				x	265	26

$\lambda_k^\delta = \lambda_k^B$, needing only slightly more iterations but significantly less computation time. As was also the case for the choice $\lambda_k^\delta = \lambda_k^B$, the choice $\lambda_k^\delta = \lambda_k^E$ behaves badly in combination with the minimal error stepsize α_k^{ME}. Again the most likely reason is the non-monotone nature of this stepsize choice.

Since the acceleration effect is due to λ_k^δ, it makes sense to look at its evolution over the course of the iteration. The left figure in figure 3 depicts the development of λ_k^E, λ_k^B and λ_k^N when used in the TPG method with steepest descent stepsize α_k^{SD} for the SPECT problem considered above. One can see that in all three cases λ_k^δ goes to one as the iteration progresses, which is the reason for the acceleration effect. Although seemingly going to one with growing k, λ_k^B stays zero for some of the first iterations and then exhibits a steep jump followed by some small oscillations, before starting to increase monotonously. This can be explained by the backtracking search procedure of the DBTS algorithm, which first has to go through some unsuccessful search cycles before the function $q(m)$ has decreased to the right order of magnitude. Afterwards, a monotonous increase also of λ_k^B can be seen. A similar phenomenon can also be observed when the DBTS algorithm is applied to the TPG method with constant stepsize ω. In the first iterations, λ_k^B is zero, then switches between zero and one before it changes to monotonous increase starting from some value in $[0, 1]$, after which it again drops to some value in $[0, 1]$ and stars yet again to increase monotonously. In combination with the minimal error stepsize, λ_k^B first exhibits the same pattern as with the steepest descent stepsize α_k^{SD} but, after a certain amount of increase, starts to decreases monotonously, which explains why the acceleration effect is lost.

Note that if the function q is chosen such that it decreases too fast, then λ_k^B will become a decreasing sequence. For example, the function $q(m) = 1/2^m$ often led to a decreasing sequence λ_k^B in our experiments. Hence, in order to profit from an acceleration effect, one has to choose a slowly decreasing function satisfying (3.22), like $q(m) = 1/m^{1+\alpha}$ with a small $\alpha > 0$. Similar restrictions can also be observed for second order methods like the Levenberg–Marquardt or the iteratively regularized Gauss–Newton method.

The right figure in figure 3 depicts the development of the norm of the residuals during the iterations of the TPG methods using the steepest descent stepsize α_k^{SD} together with the different choices of λ_k^δ considered above. Once again, one can clearly see the

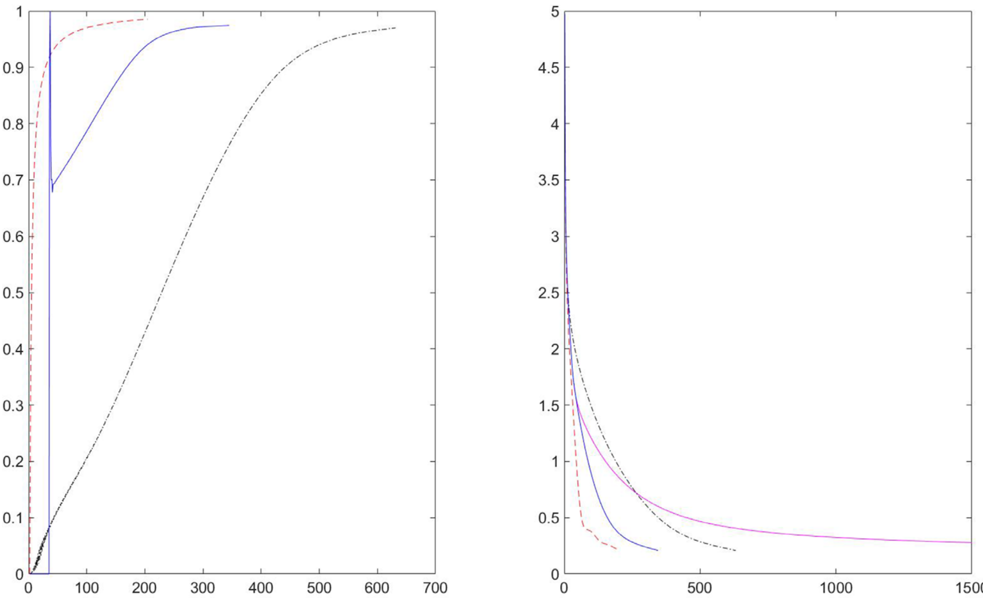

Figure 3. Results of using the TPG methods with steepest descent stepsize α_k^{SD} and various choices of λ_k^δ, using a relative data error of 0.25%. Left: plot of the values of λ_k over the iteration number k. Right: plot of the residual $\left\|A(f_k,\mu_k)-y^\delta\right\|$ over the iteration number k. Dashed red line: $\lambda_k^\delta=\lambda_k^N$, solid blue line: $\lambda_k^\delta=\lambda_k^B$, dash-dotted black line: $\lambda_k^\delta=\lambda_k^E$, solid magenta line in the right figure (extending up to the y-axis value 1500): $\lambda_k^\delta=0$.

acceleration effect due to the three considered parameters λ_k^E, λ_k^B and λ_k^N, which manage to decrease the residual norm much faster than in the case when no acceleration, i.e. $\lambda_k^\delta=0$, is being used.

Note that the residual norms decrease monotonously, which is also the case for the other stepsizes, except for the case when the minimal error stepsize α_k^{ME} is used in combination with either $\lambda_k^\delta=\lambda_k^E$ or $\lambda_k^\delta=0$, in which case oscillations occur.

In figure 4, one can see the results of the reconstruction of the activity and the attenuation function achieved when using the TPG method with steepest descent stepsize α_k^{SD} combined with λ_k^B for the choices of parameters as above and with a relative data error $\delta=0.25\%$. One can see that the activity function f_* is nicely reconstructed. The attenuation function, however, does not resemble the true attenuation function μ_* at all. This phenomenon is common for SPECT and has already been observed in [23]. The reason for this is the high nonlinearity of the problem, leading to non-uniqueness of the solution and therefore, since the reconstruction algorithm selects a solution with minimal distance to $(f_0,\mu_0)=(0,0)$, to the reconstruction of μ_* as seen in figure 4. Possible remedies already mentioned in [23] are for example a better initial guess or a coupled tomography approach. In any case, the main reason for including μ in the reconstruction is to arrive at reconstructions conforming to the data. Besides, this paper did not aim at improving the reconstruction quality of SPECT, but at showing the acceleration effect of TPG methods of the form (1.17).

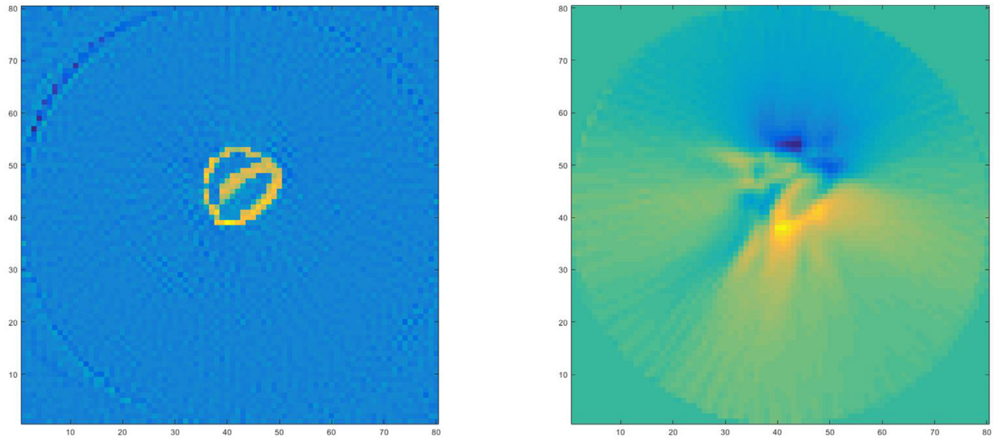

Figure 4. Results of the TPG method using the steepest descent stepsize α_k^{SD} together with $\lambda_k^\delta = \lambda_k^B$ for the SPECT example problem with a relative data error of $\delta = 0.25\%$. Activity function f_{k_*} (left) and attenuation function μ_{k_*} (right).

5. Conclusion and outlook

We have proven convergence of general TPG methods of the form (1.17) under classical assumptions for iterative regularization methods for nonlinear ill-posed problems. Afterwards, we have applied the theory to various TPG methods using either the steepest descent, the minimal error or a constant stepsize, together with different choices for the combination parameters λ_k^δ.

Although no analytical results are yet available proving that indeed less iterations are required when using TPG methods (1.17), the numerical simulation results presented above clearly show their advantages in practise. Besides the fact that much fewer iterations are necessary to arrive at suitable solutions, the implementation of TPG methods is exceedingly simple. Furthermore, they require hardly more computation time than their non-accelerated counterparts. Due to the numerically demonstrated great reduction of the required number of iterations, TPG methods could serve as a viable alternative to commonly used 'fast' iterative methods like the iteratively regularized Gauss–Newton method, especially when dealing with large-scale inverse problems, where the latter ones often become impracticable due to having to solve huge and usually full linear systems in each iteration step.

As a final comment, note that the TPG method (1.17) can also be rewritten in terms of z_k^δ, leading to

$$\begin{aligned} z_{k+1}^\delta &= (1+\lambda_{k+1}^\delta)(z_k^\delta + \alpha_k^\delta s_k^\delta) - \lambda_{k+1}^\delta(z_{k-1}^\delta + \alpha_{k-1}^\delta s_{k-1}^\delta) \\ &= z_k^\delta + \lambda_{k+1}^\delta(z_k^\delta - z_{k-1}^\delta) + (1+\lambda_{k+1}^\delta)\alpha_k^\delta s_k^\delta - \lambda_{k+1}^\delta \alpha_{k-1}^\delta s_{k-1}^\delta, \end{aligned} \tag{5.1}$$

and it therefore structurally differs from the iteration methods considered by Scherzer in [25] by the additional term $\lambda_{k+1}^\delta(z_k^\delta - z_{k-1}^\delta)$. However, many of his ideas and arguments for proving convergence of those methods were re-used in the proofs of this paper.

6. Support and acknowledgements

The authors were partly funded by the Austrian Science Fund (FWF): W1214-N15, project DK8. Furthermore, they would like to thank Dr Gerhard Hubmer, Dr Stefan Kindermann and

Prof Dr Andreas Neubauer for reading early drafts of the manuscript and providing valuable suggestions and insights during discussions of the subject.

References

[1] Attouch H and Peypouquet J 2016 The rate of convergence of Nesterov's accelerated forward–backward method is actually faster than $O(1/k^2)$ *SIAM J. Optim.* **26** 1824–34
[2] Beck A and Teboulle M 2009 A fast iterative shrinkage-thresholding algorithm for linear inverse problems *SIAM J. Imaging Sci.* **2** 183–202
[3] Blaschke B, Neubauer A and Scherzer O 1997 On convergence rates for the iteratively regularized Gauss–Newton method *IMA J. Numer. Anal.* **17** 421
[4] Dicken V 1998 Simultaneous activity and attenuation reconstruction in single photon emission computed tomography, a nonlinear ill-posed problem *PhD Thesis* Universität Potsdam
[5] Dicken V 1999 A new approach towards simultaneous activity and attenuation reconstruction in emission tomography *Inverse Problems* **15** 931
[6] Engl H W, Hanke M and Neubauer A 1996 *Regularization of Inverse Problems* (Dordrecht: Kluwer)
[7] Hanke M 1991 Accelerated Landweber iterations for the solution of ill-posed equations *Numer. Math.* **60** 341–73
[8] Hanke M 1997 A regularizing Levenberg–Marquardt scheme, with applications to inverse groundwater filtration problems *Inverse Problems* **13** 79
[9] Hanke M, Neubauer A and Scherzer O 1995 A convergence analysis of the Landweber iteration fornonlinear ill-posed problems *Numer. Math.* **72** 21–37
[10] Terry J A, Tsui B M W, Hendricks J R and Gullberg G T 1990 The design of a mathematical phantom of the upper human torso for use in 3D SPECT imaging research *Proc.* 1990 *Fall Mtg of the Biomedical Engineering Society* (*Blacksburg, VA*) *pp* 1467–74
[11] Jin Q 2010 On a regularized Levenberg–Marquardt method for solving nonlinear inverse problems *Numer. Math.* **115** 229–59
[12] Jin Q 2016 Landweber–Kaczmarz method in Banach spaces with inexact inner solvers *Inverse Problems* **32** 104005
[13] Jin Q and Tautenhahn U 2009 On the discrepancy principle for some Newton type methods for solving nonlinear inverse problems *Numer. Math.* **111** 509–58
[14] Kaltenbacher B, Neubauer A and Scherzer O 2008 *Iterative Regularization Methods for Nonlinear Ill-Posed Problems* (Berlin: de Gruyter)
[15] Natterer F 2001 *The Mathematics of Computerized Tomography* (Philadelphia: SIAM) (https://doi.org/10.1137/1.9780898719284)
[16] Nesterov Y 1983 A method of solving a convex programming problem with convergence rate $O(1/k^2)$ *Sov. Math. Dokl.* **27** 372–6
[17] Neubauer A 2000 On Landweber iteration for nonlinear ill-posed problems in Hilbert scales *Numer. Math.* **85** 309–28
[18] Neubauer A 2016 Some generalizations for Landweber iteration for nonlinear ill-posed problems in Hilbert scales *J. Inverse Ill-Posed Problems* **24** 393–406
[19] Neubauer A 2017 A new gradient method for ill-posed problems submitted
[20] Neubauer A 2017 On Nesterov acceleration for Landweber iteration of linear ill-posed problems *J. Inverse Ill-Posed Problems* **25** 381–90
[21] Ramlau R 1999 A modified landweber method for inverse problems *Numer. Funct. Anal. Optim.* **20** 79–98
[22] Ramlau R 2003 TIGRA—an iterative algorithm for regularizing nonlinear ill-posed problems *Inverse Problems* **19** 433
[23] Ramlau R and Teschke G 2006 A Tikhonov-based projection iteration for nonlinear ill-posed problems with sparsity constraints *Numer. Math.* **104** 177–203
[24] Hubmer S, Neubauer A, Ramlau R and Voss H U 2016 On the parameter estimation problem of magnetic resonance advection imaging *Inverse Problems Imaging* submitted
[25] Scherzer O 1996 A convergence analysis of a method of steepest descent and a twostep algorothm for nonlinear ill-posed problems *Numer. Funct. Anal. Optim.* **17** 197–214