



# Trilevel and multilevel optimization using monotone operator theory

Allahkaram Shafiei<sup>1</sup> · Vyacheslav Kungurtsev<sup>1</sup> · Jakub Marecek<sup>1</sup>

Received: 2 October 2022 / Revised: 15 January 2024 / Accepted: 16 January 2024 /

Published online: 28 February 2024

© The Author(s) 2024

## Abstract

We consider rather a general class of multi-level optimization problems, where a convex objective function is to be minimized subject to constraints of optimality of nested convex optimization problems. As a special case, we consider a trilevel optimization problem, where the objective of the two lower layers consists of a sum of a smooth and a non-smooth term. Based on fixed-point theory and related arguments, we present a natural first-order algorithm and analyze its convergence and rates of convergence in several regimes of parameters.

**Keywords** Variational inequality · Bi-level optimization · Tri-level optimization · Multi-level minimization · Non-expansive mappings

## 1 Introduction

Hierarchical Optimization Problems, also known as Multilevel Optimization Problems (MOP), were first introduced by Bialas and Karwan (1973) and Wilfred (1977) as a class of constrained optimization problems, wherein the feasible set is determined—implicitly—as the optima of multiple optimization problems, nested in a predetermined sequence. In theory, MOP has applications in game theory, robust optimization, chance-constrained programming, and adversarial machine learning (Ben-Tal et al. 2009). In practice, MOP models are widely used in security applica-

---

✉ Jakub Marecek  
jakub.marecek@fel.cvut.cz

Allahkaram Shafiei  
shafiall@fel.cvut.cz

Vyacheslav Kungurtsev  
vyacheslav.kungurtsev@fel.cvut.cz

<sup>1</sup> Department of Computer Science, Czech Technical University, Karlovo Namesti 13, 121 35 Prague 2, Czech Republic

tions, where they model so-called interdiction problems. See Iiduka (2011); Moudafi (2007); Xu (2010); Isao (2001) for several examples.

Tri-level programming problems are challenging, even when one considers continuous linear problems (Blair 1992) due to their computational complexity and the interactions between decision-makers based on the numbers of variables at different levels. Moreover, at each level, we have limited or incomplete information about the decisions made at the other levels. This interdependency makes it difficult to decouple the optimization problem into separate sub-problems. This, in turn, requires techniques that can handle the hierarchical nature of the problem.

The article is structured as follows. Section 1 introduces the necessary notations, assumptions, and the problem model. Additionally, it establishes fundamental notation and background information pertaining to the proximal-gradient algorithm. Section 2 delves into convergence analysis for a variety of assumptions on the step sizes. Furthermore, in this section, we present the methodological approach employed in our study regarding error bounds that enables us to provide the convergence of sequence generated by the algorithm for tri-level problems. In Sect. 3, the paper introduces the convergence rate analysis for variants of the proximal-gradient algorithm. It shows, among others, that our convergence rate in the middle layer, which is  $\mathcal{O}(\frac{1}{(k+1)\sqrt{k}})$ , improves upon the rate of Sabach and Shtern (2017). Finally, in Sect. 4, we formalize multi-level optimization, followed by an exploration of both convergence and the corresponding convergence rate.

## 1.1 The problem

In particular, our goal is to formulate and analyze an optimization algorithm for a class of hierarchically-defined problems:

$$\begin{cases} \min_{x \in X_N^*} \omega(x) \\ X_i^* = \arg \min_{x \in X_{i-1}^*} [f_i(x) + g_i(x)], \quad i \in \{1, \dots, N\} \\ X_0^* = \mathbb{R}^n, \end{cases} \quad (1)$$

where the middle layers (for  $i \in \{1, \dots, N\}$ ) exhibit the so-called composite structure, where  $\omega$  is a strongly convex differentiable function and there are smooth terms  $f_i$  and non-smooth terms  $g_i$ . In machine-learning applications, the smooth functions are chosen to be loss functions and the non-smooth functions  $g_i$  are regularizers.

We begin by considering a hierarchical optimization with three layers, wherein the middle and lower layers exhibit the composite structure:

$$\begin{cases} \min_{x \in X^*} \omega(x) \\ X^* = \arg \min_{x \in Y^*} [\phi_2(x) := f_2(x) + g_2(x)] \\ Y^* = \arg \min_{x \in \mathbb{R}^n} [\phi_1(x) := f_1(x) + g_1(x)]. \end{cases} \quad (2)$$

By leveraging fixed-point theory and related reasoning, we propose a straightforward first-order algorithm and analyze its convergence and convergence rates across various parameter regimes. The algorithm exhibits the following non-asymptotic behaviour: The first layer exhibits a convergence rate of  $\mathcal{O}(\frac{1}{k})$ , the second layer (middle layer) exhibits a convergence rate  $\mathcal{O}(\frac{1}{(k+1)\sqrt{k}})$ , and finally, the third layer exhibits  $\mathcal{O}(\frac{1}{\sqrt{k}})$  global rate of convergence concerning the inner objective function values. By accessing the main iteration in terms of the inner objective function values, we observe the convergence rate  $\mathcal{O}(\frac{1}{k})$ .

## 1.2 Related work

Our work is inspired by a long history of work on bilevel optimization problems (see, e.g., Al-Khayyal et al. (1992); Dempe et al. (2007, 2014); Zhang (1994)). Our work extends proximal-gradient optimization algorithms (Sabach and Shtern 2017) for a related bilevel optimization problem and is informed by Moudafi (2007).

Notably, Solodov (2007) gave an explicit descent method for bi-level optimization in the form of

$$\begin{cases} \min \omega(x) \\ x \in S := \arg \min_{x \in F} f(x) = \arg \min_{x \in \mathbb{R}^n} [f(x) + i_F], \end{cases} \quad (3)$$

in which  $i_F$  is the indicator function on  $F$  and  $f$  is convex and smooth function and  $\omega$  is strongly convex. Subsequently, Sabach and Shtern (2017) proposed and Lampariello et al. (2020) further developed the so-called BIG-SAM method for solving the more general problem of,

$$\begin{cases} \min \omega(x) \\ x \in Y^* := \arg \min_{x \in \mathbb{R}^n} [f(x) + g(x)], \end{cases} \quad (4)$$

in which  $f$  is smooth and  $g$  is convex and lower semi-continuous and possibly non-smooth. We consider a similar structure in a multi-level problem.

There are only a few solution approaches presented in the literature for tri-level problems, addressing very restricted classes of problems, and mostly without guarantees of global optimality. For example, error-bound conditions were used by Senter and Dotson (1974) to assure the existence of strong convergence results for Mann iterates. Typically, error bounds are essential for assessing the accuracy and reliability of numerical approximations or algorithms and, providing a measure of how close the approximate solution is to the true solution, given certain assumptions or conditions. These conditions may include properties of the problem, the algorithm used, the precision of numerical calculations, and any assumptions made during the approximation process. Very recently, Sato et al. (2021) presented a gradient-based algorithm for multilevel optimization, where the lower-level problems are replaced by steepest descent update equations. They present conditions when this reformulation

asymptotically converges to the original multilevel problem. Based on our knowledge, no other solution approach can tackle the class of problems considered in this work.

### 1.3 Examples

We present several concrete examples of trilevel optimization problems.

1. Pursuit-evasion-intercept, or alternatively described as pursuit-evade-defend (see, e.g. Fisac et al. 2015) is a (sequential) game wherein one player is seeking to follow and capture another in a dynamic setting, with a third tasked with intercepting the pursuer or
2. Bilevel optimization with robust uncertainty. Robust optimization, i.e., choosing the optimal outcome upon the worst case realization of a parameter. This can be expressed as a nested optimization problem, wherein the inner problem is a maximum over the parameter set (Lampariello et al. 2019). Any classic bilevel optimization, for instance Stackelberg games, can become trilevel when the leader makes a decision under robust uncertainty consideration.
3. Mixture models with training and validation: consider some convex loss function on data with a regularization (e.g., LASSO), wherein the validation (for instance, a coresets) data set is considered more significant and thus an inner problem, the training set presents the middle problem, and the tuning of mixture weights of different models is the uppermost layer.

### 1.4 Preliminaries

Let  $\Omega \subseteq \mathbb{R}^n$  be closed and convex and let  $T$  be a mapping from  $\mathbb{R}^n$  into itself. Recall that the notion of the variational inequality (VI), denoted by  $VI(T, \Omega)$ , is to find a vector  $x^* \in \Omega$  such that

$$VI(T, \Omega) \quad \langle x^* - T(x^*), x - x^* \rangle \geq 0 \quad \forall x \in \Omega. \quad (5)$$

Note that (5) is equivalent to finding the fixed point of the problem

$$\text{Find } x^* \in \Omega \text{ such that } x^* = P_{\Omega}T(x^*),$$

where  $P_{\Omega}$  is the metric projection of  $\mathbb{R}^n$  onto  $\Omega$ , i.e., it maps  $x \in \mathbb{R}^n$  to the unique point in  $\Omega$  defined as, where throughout the paper we use the Euclidean norm,

$$P_{\Omega}(x) := \arg \min_{y \in \Omega} \|x - y\| \quad \forall x \in \mathbb{R}^n,$$

satisfying

$$\|x - P_{\Omega}(x)\| = \inf \{\|x - y\| : y \in \Omega\} := d(x, \Omega), \quad P_{\Omega}(x) \in \Omega.$$

We also use the notation  $\text{Fix}(T) = \{x \in \mathbb{R}^n : T(x) = x\}$  for the set of fixed points of  $T$ . This set  $\text{Fix}(T)$  is closed and convex for non-expansive mappings  $T$ .

We now review some well-known facts about non-expansive mappings that we shall henceforth use in the paper without reference.

- Let  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$  be a non-expansive mapping. Then  $I - T$  is monotone; that is

$$\langle x - y, (I - T)x - (I - T)y \rangle \geq 0.$$

- Let  $S : \mathbb{R}^n \rightarrow \mathbb{R}^n$  be a contraction mapping with coefficient  $r \in (0, 1)$ . Then  $I - S$  called  $(1 - r)$ -strongly monotone; that is

$$\langle x - y, (I - S)x - (I - S)y \rangle \geq (1 - r)\|x - y\|^2, \quad \forall x, y \in \mathbb{R}^n.$$

- Let  $x \in \mathbb{R}^n$  and  $z \in \Omega$  be given. Then  $z = P_\Omega(x)$  if and only if the following inequality holds

$$\langle x - z, z - y \rangle \geq 0 \quad \forall y \in \Omega,$$

and also if and only if

$$\|x - z\|^2 + \|y - z\|^2 \leq \|x - y\|^2, \quad \forall y \in \Omega.$$

- For all  $x, y \in \mathbb{R}^n$  one has

$$\|P_\Omega(x) - P_\Omega(y)\|^2 \leq \langle P_\Omega(x) - P_\Omega(y), x - y \rangle.$$

## 2 Convergence analysis for trilevel optimization problems

Let us consider the trilevel problem (2). After presenting our assumptions and preliminaries, we analyze the convergence of a proximal-gradient algorithm first under a variety of conditions on the step sizes (Sect. 2.3, using lemmas from Sect. 2.2). Alternatively, one can assume a certain error-bound condition (Sect. 2.4).

### 2.1 Assumptions and the algorithm

We shall make the following standing assumptions:

**Assumption 1** (i)  $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$  are convex and continuously differentiable with  $L_{f_i}$ -Lipschitz gradient, that is,

$$\|\nabla f_i(x) - \nabla f_i(y)\| \leq L_{f_i}\|x - y\|, \quad \forall x, y \in \mathbb{R}^n, \quad i = 1, 2.$$

- (ii)  $g_i : \mathbb{R}^n \rightarrow (-\infty, +\infty]$  is proper, lower semi-continuous and convex.
- (iii) the optimal solution set of the inner layers is non-empty, i.e.,  $X^* \neq \emptyset$  and  $Y^* \neq \emptyset$ .
- (iv)  $\omega : \mathbb{R}^n \rightarrow \mathbb{R}$  is strongly convex with strong convexity parameter  $\mu$ .
- (v)  $\omega$  is a continuously differentiable function so that  $\nabla \omega(\cdot)$  is Lipschitz continuous with constant  $L_\omega$ .

**Table 1** The three layers of a trilevel problem and the corresponding operators

Layer	Function	Operator	Solution set	Optimality condition
Top	$\omega(x)$	$S$		$\langle x^* - S(x^*), x - x^* \rangle \geq 0, \forall x \in X^*$
Middle	$f_2(x) + g_2(x)$	$T$	$X^* \subseteq Y^*$	$\langle x^* - T(x^*), x - x^* \rangle \geq 0, \forall x \in \text{Fix}(W) = Y^*$
Bottom	$f_1(x) + g_1(x)$	$W$	$Y^*$	$x^* \in \text{Fix}(W)$

Consider three operators corresponding to the three layers of objective functions,

$$\begin{aligned} S(x) &:= S_u(x) = x - u \nabla \omega(x), \\ T(x) &:= T_t(x) = \text{prox}_{tg_2}(x - t \nabla f_2(x)), \\ W(x) &:= W_s(x) = \text{prox}_{sg_1}(x - s \nabla f_1(x)). \end{aligned} \quad (6)$$

Note that each of these corresponds to a fixed point map of their respective problems. It can be easily seen that  $X^* \cap \text{Fix}(T) = \text{Fix}(T) \cap \text{Fix}(W)$ , however, in general, we are only interested in a (specific subset) of  $\text{Fix}(W) = Y^*$  and we expect  $\text{Fix}(T) \cap \text{Fix}(W)$  to be empty.

It is well known that mappings  $T$  and  $W$  are non-expansive and  $S$  is an  $r$ -contraction, i.e., for any  $u \in (0, \frac{2}{L_\omega + \mu}]$ , one has

$$\|S(x) - S(y)\| \leq \sqrt{1 - \frac{2u\mu L_\omega}{\mu + L_\omega}} \|x - y\|, \quad (7)$$

[For more details, see (Nesterov 2003, Theorem 2.1.12, p. 66).

For any proper, lower semi-continuous and convex function  $g : \mathbb{R}^n \rightarrow (-\infty, +\infty]$  the Moreau proximal mapping is defined by

$$\text{prox}_g(x) = \arg \min_{u \in \mathbb{R}^n} \left\{ g(u) + \frac{1}{2} \|u - x\|^2 \right\}. \quad (8)$$

In general, a proximal-gradient algorithm (Beck and Sabach 2014) is based on an iterated mapping:

$$T_t(x) = \text{prox}_{tg}(x - t \nabla f(x)),$$

which has the following properties:

(i)  $T_t$  is non-expansive for sufficiently small  $t$ , i.e.,

$$\|T_t(x) - T_t(y)\| \leq \|x - y\| \quad \forall x, y \in \mathbb{R}^n, \quad \forall t \in \left(0, \frac{1}{L_f}\right]. \quad (9)$$

(ii) Its fixed points are equivalent to the set of minimizers to the corresponding minimization problem, i.e.,

$$\text{Fix}(T_t) = \arg \min_{x \in \mathbb{R}^n} [f(x) + g(x)] \quad \forall t > 0. \quad (10)$$

We shall denote the proximal gradient mapping as  $T$  in the Algorithm, instead of  $T_t$ , because of property (ii).

A solution  $x^*$  of (2) satisfies the following inequalities,

$$\begin{aligned} &\text{Find } x^* \in Y^* = \text{Fix}(W) \\ &\langle x^* - T(x^*), x - x^* \rangle \geq 0, \quad \forall x \in \text{Fix}(W) = Y^* \\ &\langle x^* - S(x^*), x - x^* \rangle \geq 0, \quad \forall x \in X^*. \end{aligned} \quad (11)$$

Note that  $x \in X^*$  is equivalent to  $\langle x^* - T(x^*), x - x^* \rangle \geq 0, \forall x \in \text{Fix}(W) = Y^*$  so the third condition can be modified to:  $\langle x^* - S(x^*), x - x^* \rangle \geq 0$ , for all  $x$  satisfying this relation.

For reference regarding problems (2) and (4), we present Table 1.

Throughout this section, we are concerned with Algorithm 1, which is based on the proximal-gradient maps and their following combination:

$$x^{k+1} = \alpha_k S(x^k) + (1 - \alpha_k) \beta_k T(x^k) + (1 - \alpha_k)(1 - \beta_k) W(x^k). \quad (12)$$

Define the following quantities regarding the relative limit behaviors of the two parameters:

$$\delta := \limsup_{k \rightarrow \infty} \frac{\beta_k}{\alpha_k} \in [0, \infty] \text{ and } \tilde{\delta} := \lim_{k \rightarrow \infty} \frac{\beta_k}{\alpha_k} \in [0, \infty]. \quad (13)$$

These quantities play a central role in analyzing the convergence of the proposed algorithms. For more details, you may see Example 4.

The following key Assumptions will be needed throughout the paper:

**Assumption 2**  $\alpha_k \rightarrow 0$  (as  $k \rightarrow \infty$ ) and  $\sum_{k=1}^{+\infty} \alpha_k = \infty$ .

**Assumption 3** There exists  $K > 0$  such that  $\limsup_{k \rightarrow \infty} \frac{1}{\alpha_k} \left| \frac{1}{\beta_k} - \frac{1}{\beta_{k-1}} \right| \leq K$ .

**Assumption 4**  $\limsup_{k \rightarrow \infty} \frac{|\beta_k - \beta_{k-1}| + |\alpha_k - \alpha_{k-1}|}{\alpha_k \beta_k} = 0$ .

## 2.2 Properties of the limit points

We now derive a set of results regarding the properties of limit points generated by the sequence (12). First, we present the following powerful lemma that we shall use in the analysis below:

**Algorithm 1** Proximal Gradient for Tri-level Optimization

(t!) **Input:**  $t \in \left(0, \frac{1}{L_{f_1}}\right]$ ,  $s \in \left(0, \frac{1}{L_{f_2}}\right]$ ,  $r \in \left(0, \frac{2}{L_{\omega} + \mu}\right]$  and the real sequences  $\alpha_k$  and  $\beta_k$  satisfy the assumptions

**Initialization:** Select an arbitrary starting point  $x^0 \in \mathbb{R}^n$

**For**  $k = 1, 2, \dots$  **do**

$$\begin{aligned} y^k &:= \text{prox}_{t g_2} \left[ x^{k-1} - t \nabla f_2(x^{k-1}) \right] \\ z^k &:= \text{prox}_{s g_1} \left[ x^{k-1} - s \nabla f_1(x^{k-1}) \right] \\ v^k &:= x^{k-1} - r \nabla \omega(x^{k-1}) \\ x^k &= \alpha_k v^k + (1 - \alpha_k) \beta_k y^k + (1 - \beta_k)(1 - \alpha_k) z^k \end{aligned}$$

**End For**

**Lemma 5** (Xu 2002, Lemma 2.1) Assume that  $a_k$  be a sequence of non-negative real numbers such that

$$a_{k+1} \leq (1 - \gamma_k) a_k + \delta_k,$$

where  $\gamma_k$  is a sequences in  $(0, 1)$  and  $\delta_k$  is a sequence in  $\mathbb{R}$ , such that

- (1)  $\sum_{k=1}^{\infty} \gamma_k = \infty$ ,  
 (2) either  $\limsup_{k \rightarrow \infty} \frac{\delta_k}{\gamma_k} \leq 0$  or  $\sum_{k=1}^{\infty} |\delta_k| < \infty$ .

Then  $\lim_{k \rightarrow \infty} a_k = 0$ .

From now on and throughout the paper, we denote by  $\{x^k\}$  the sequence generated by the algorithm (12). The convergence of the algorithm crucially depends on the starting points  $x^0 \in \mathbb{R}^n$  and the parameters (step-sizes)  $\alpha_k$  and  $\beta_k$ , which are chosen in advance. Three different cases can be distinguished:  $\delta = 0$ ,  $\delta > 0$  and  $\delta = \infty$ , each associated with some other Assumptions. Initially, we are going to seek the conditions ensuring boundedness of the sequence of iterates  $\{x^k\}$ . Our proof techniques are similar to those that (Sabach and Shtern 2017) used to prove their Lemma 2.

Throughout this paper, to simplify the notation, we will use  $w(\{x^k\})$  to denote the set of cluster points of sequence  $\{x^k\}$ , i.e.,

$$w(\{x^k\}) = \left\{ x \in \mathbb{R}^n : x^{k_i} \rightarrow x \text{ for some sub-sequence } \{x^{k_i}\} \text{ of } \{x^k\} \right\},$$

and also for every  $k \geq 1$ , we define

$$Q_k(x) = \beta_k T(x) + (1 - \beta_k) W(x).$$

It is straightforward to see that  $Q_k$  is non-expansive.

**Lemma 6** Assume  $\delta < \infty$ . Then  $\{x^k\}$  is bounded, i.e., for every  $x \in \text{Fix}(W)$  there exists a constant  $C_x$  such that  $\|x^k - x\| \leq C_x$  and constants  $C_S$  and  $C_T$  such that

$$\|W(x^k) - x\| \leq C_x, \quad \|S(x^k) - x\| \leq C_S + C_x, \quad \|T(x^k) - x\| \leq C_T + C_x.$$



Moreover, for all  $x \in \text{Fix}(W)$  one has

$$\limsup_k (\|x^{k+1} - x\| - \|x^k - x\|) \leq 0.$$

**Proof** Taking into account  $\delta \in [0, +\infty)$ , from Assumption (13) one sees that there exists  $\delta_0 > \delta$  and  $k_0 \in \mathbb{N}$  such that for every  $k \geq k_0$ , one has  $\beta_k < \delta_0 \alpha_k$ . On the other hand the sequence  $\{x^{k+1}\}$  can easily be rewritten as

$$x^{k+1} = \alpha_k S(x^k) + (1 - \alpha_k) Q_k(x^k).$$

Now, for given  $x \in \text{Fix}(W)$  we obtain

$$\begin{aligned} \|x^{k+1} - x\| &= \|\alpha_k(S(x^k) - S(x)) + \alpha_k(S(x) - x) \\ &\quad + (1 - \alpha_k)(Q_k(x^k) - Q_k(x)) + (1 - \alpha_k)(Q_k(x) - x)\| \\ &\leq \alpha_k r \|x^k - x\| + \alpha_k \|S(x) - x\| \\ &\quad + (1 - \alpha_k) \|x^k - x\| + (1 - \alpha_k) \beta_k \|T(x) - x\| \\ &\leq (1 - (1 - r)\alpha_k) \|x^k - x\| + \alpha_k (\|S(x) - x\| + \delta_0 \|T(x) - x\|) \\ &\leq \max \left\{ \|x^k - x\|, \frac{1}{1 - r} (\|S(x) - x\| + \delta_0 \|T(x) - x\|) \right\} \\ &\leq \max \left\{ \|x^{k-1} - x\|, \frac{1}{1 - r} (\|S(x) - x\| + \delta_0 \|T(x) - x\|) \right\} \\ &\leq \dots \leq \max \left\{ \|x^{k_0} - x\|, \frac{1}{1 - r} (\|S(x) - x\| + \delta_0 \|T(x) - x\|) \right\} := C_x. \end{aligned} \quad (14)$$

There,  $r$  is the coefficient of the contraction map  $S$ . Therefore,  $\{x^k\}$  is bounded. Also, for given  $x \in \text{Fix}(W)$  from (14) one can observe that

$$\|x^{k+1} - x\| \leq (1 - (1 - r)\alpha_k) \|x^k - x\| + \alpha_k \|S(x) - x\| + (1 - \alpha_k) \beta_k \|T(x) - x\|, \quad (15)$$

which implies that

$$\limsup_k (\|x^{k+1} - x\| - \|x^k - x\|) \leq 0.$$

□

**Remark 7** One can see that if  $\text{Fix}(W) \cap \text{Fix}(T) \neq \emptyset$ , then  $\{x^k\}$  is bounded, without taking into consideration the condition  $\delta \in [0, \infty]$ . Indeed when  $x \in \text{Fix}(W) \cap \text{Fix}(T)$

by (15) we then have

$$\begin{aligned}\|x^{k+1} - x\| &\leq (1 - (1 - r)\alpha_k)\|x^k - x\| + \alpha_k\|S(x) - x\| \\ &\leq \max\{\|x^k - x\|, \frac{\|S(x) - x\|}{1 - r}\} \\ &\leq \dots \leq \max\{\|x^{k_0} - x\|, \frac{\|S(x) - x\|}{1 - r}\},\end{aligned}$$

which shows that  $\{x^k\}$  is bounded.

The following simple example shows that the boundedness of  $\{x^k\}$  does not necessarily hold when  $\delta = \infty$ .

**Example 1** Take  $X = \mathbb{R}$  and  $\alpha_k = \frac{1}{k}$  and  $\beta_k = \frac{1}{\sqrt{k}}$ . Furthermore, let  $S(x) = \frac{x}{4}$ ,  $T(x) = x + 5$ ,  $W(x) = x$ . Clearly,  $S$  is contraction, and  $T$  and  $W$  are non-expansive. It is easy to check that  $\delta = \limsup_k \frac{\beta_k}{\alpha_k} = \infty$ , and  $x^k \rightarrow \infty$  from starting point  $x^0 = 1$ .

The next Lemma will be useful in the sequel the proof is slightly similar to (Lu et al. 2009, Theorem 4.1)

**Lemma 8** Suppose that  $\{x^k\}$  is bounded.

(a) If Assumption 4 holds, then  $\{x^k\}$  is asymptotically regular, i.e.,

$$\lim_{k \rightarrow \infty} \|x^{k+1} - x^k\| = 0.$$

(b) If Assumptions 2, 3 and 4 hold, then  $\lim_{k \rightarrow \infty} \left\| \frac{x^{k+1} - x^k}{\beta_k} \right\| = 0$ .

(c) one has  $w(\{x^k\}) \subseteq \text{Fix}(W)$ .

**Proof** Since  $\{x^k\}$  is bounded then there exists constant  $M$  such that

$$M \geq \sup_{k \geq 1} \left\{ \|S(x^{k-1})\|, \|T(x^{k-1})\|, \|W(x^{k-1})\| \right\}.$$

So we have

$$\begin{aligned}\|Q_k(x^k) - Q_{k-1}(x^{k-1})\| &= \|Q_k(x^k) - Q_k(x^{k-1}) + Q_k(x^{k-1}) - Q_{k-1}(x^{k-1})\| \\ &= \|Q_k(x^k) - Q_k(x^{k-1}) + (\beta_k - \beta_{k-1})(T(x^{k-1}) - W(x^{k-1}))\| \\ &\leq \|x^k - x^{k-1}\| + 2M|\beta_k - \beta_{k-1}|.\end{aligned}$$

Now, one can write

$$\begin{aligned}
 \|x^{k+1} - x^k\| &= \|(\alpha_k S(x^k) \\
 &\quad + (1 - \alpha_k) Q_k(x^k) - (\alpha_{k-1} S(x^{k-1}) + (1 - \alpha_{k-1}) Q_{k-1}(x^{k-1}))\| \\
 &= \|(1 - \alpha_k)(Q_k(x^k) - Q_{k-1}(x^{k-1})) + (\alpha_k - \alpha_{k-1})(S(x^{k-1}) - Q_{k-1}(x^{k-1})) \\
 &\quad + \alpha_k(S(x^k) - S(x^{k-1}))\| \\
 &\leq (1 - (1 - r)\alpha_k)\|x^k - x^{k-1}\| + 2M|\beta_k - \beta_{k-1}| + 2M|\alpha_k - \alpha_{k-1}| \\
 &\leq (1 - (1 - r)\alpha_k)\|x^k - x^{k-1}\| + \frac{2M|\beta_k - \beta_{k-1}| + 2M|\alpha_k - \alpha_{k-1}|}{\beta_k}. \quad (16)
 \end{aligned}$$

Noticing Assumption 4 and setting  $a_k = \|x^{k+1} - x^k\|$ ,  $\gamma_k = (1 - r)\alpha_k$ , and

$$\delta_k = \frac{2M|\beta_k - \beta_{k-1}| + 2M|\alpha_k - \alpha_{k-1}|}{\beta_k},$$

one can apply Lemma 5 and the proof of part (a) is complete.

To prove part (b): Dividing both sides of the inequality (16) by  $\beta_k$ , we obtain

$$\begin{aligned}
 \frac{\|x^{k+1} - x^k\|}{\beta_k} &\leq (1 - (1 - r)\alpha_k) \left( \frac{1}{\beta_k} + \frac{1}{\beta_{k-1}} - \frac{1}{\beta_{k-1}} \right) \|x^k - x^{k-1}\| \\
 &\quad + 2M \left( \frac{|\beta_k - \beta_{k-1}| + |\alpha_k - \alpha_{k-1}|}{\beta_k} \right) \\
 &\leq (1 - (1 - r)\alpha_k) \left( \frac{\|x^k - x^{k-1}\|}{\beta_{k-1}} \right) + \left| \frac{1}{\beta_k} - \frac{1}{\beta_{k-1}} \right| (\|x^k - x^{k-1}\|) \\
 &\quad + 2M \left( \frac{|\beta_k - \beta_{k-1}| + |\alpha_k - \alpha_{k-1}|}{\beta_k} \right).
 \end{aligned}$$

Using Assumptions 2, 3, 4 and by similar reasoning at part (a) the assertion follows from Lemma 5.

To prove part (c) : By boundedness of  $\{x^k\}$  and  $\alpha_k \rightarrow 0$  and  $\beta_k \rightarrow 0$  it is clear to see that

$$\begin{aligned}
 \|x^{k+1} - W(x^k)\| &= \|\alpha_k S(x^k) + (1 - \alpha_k)\beta_k T(x^k) \\
 &\quad + (1 - \alpha_k)(1 - \beta_k)W(x^k) - W(x^k)\| \\
 &= \|\alpha_k S(x^k) + (1 - \alpha_k)\beta_k T(x^k) + (\alpha_k\beta_k + \alpha_k + \beta_k)W(x^k)\| \\
 &\leq \alpha_k\|S(x^k)\| + (1 - \alpha_k)\beta_k\|T(x^k)\| + (\alpha_k\beta_k + \alpha_k + \beta_k)\|W(x^k)\|
 \end{aligned}$$

and we have  $\|x^{k+1} - W(x^k)\| \rightarrow 0$  which together with part (a) gives the conclusion of part (c).  $\square$

By virtue of the prior lemma, we are able, in many situations, to get a unique solution for the multilevel variational inequality without additional conditions on mappings  $S, T, W$ .

**Table 2** An overview of our results in Sects. 2.3—with the corresponding assumptions on  $\alpha_k$  and  $\beta_k$  and examples of the series

Assumptions	$\delta := \limsup \frac{\beta_k}{\alpha_k}$	$\tilde{\delta} := \lim \frac{\beta_k}{\alpha_k}$	An example	Results
Ass. 2, 11	$\delta = 0$	..	Ex. 5 (a)	Thm. 12
Ass. 2, 3, 4	$\delta \in [0, +\infty)$	..	Ex. 5 (c)	Prop. 15
Ass. 2, 3, 4, $(A_2)$ , $(A_3)$	..	$\tilde{\delta} = \infty$	Ex. 5 (d)	Thm. 19
Ass. 2, 3, 4, 11, $(A_2)$	$\delta \in [0, \infty)$	..	Ex. 5 (f)	Prop. 20

### 2.3 Convergence analysis under assumptions on step-sizes $\alpha_k$ and $\beta_k$

Next, we shall explore the convergence guarantees associated with different cases of  $\delta$ . We summarize these results, which depend on problem assumptions and parameter regimes, in Table 2. Let us consider the existence of a solution for the convex trilevel optimization problem (2). We will analyze this in multiple stages. It is worthwhile to note that the convergence behavior towards  $X^*$  is made complex by the interconnection among the three layers. On the whole, the non-expansive operators  $T$  and  $W$  do not increase the distance between any two points in the iteration for  $k$  large enough, and  $S$  contracts the distance between points in the sequence. As step size  $\alpha_k$  goes to zero, the ascendancy of the contraction mapping  $S$  diminishes, and the sequence  $\{x^k\}$  becomes dominated by the non-expansive mappings  $T$  and  $W$ . The exact convergence behavior to a specific fixed point in  $X^*$  will depend on additional properties of the individual operator  $\text{Fix}(W) = Y^*$  or its corresponding level  $\phi_1$ , such as the quadratic growth condition and linearly regular bound. First, let us consider the consistent case, i.e.,  $\text{Fix}(T) \cap \text{Fix}(W) \neq \emptyset$ , and subsequently, further cases depending on the error bound condition.

Let us now present the key technical lemma concerning the case of  $\tilde{\delta} = \infty$ , which relates to the convergence of the iteration. It establishes a connection between the set of cluster points of the sequence  $x^k$  and the solution set of the variational inequality  $VI(T, \text{Fix}(W))$ :

**Lemma 9** Assume  $\tilde{\delta} = \infty$ , together with Assumptions 2, 3, and 4. Furthermore, suppose that  $\{x^k\}$  is bounded. Then every cluster point of the sequence  $\{x^k\}$  is in  $VI(T, \text{Fix}(W))$ , i.e.,

$$w(\{x^k\}) \subseteq \{x \in \text{Fix}(W) : \langle (I - T)x, y - x \rangle \geq 0, \forall y \in \text{Fix}(W)\}.$$

**Proof** Let  $y_k = \frac{1}{\beta_k}(x^k - x^{k+1})$  and  $x \in w(\{x^k\})$  be given. It was shown that (27) holds for every  $y \in \text{Fix}(W)$ . Therefore,

$$\langle y^k, x^k - y \rangle \geq \frac{\alpha_k}{\beta_k} \langle (I - S)(x^k), x^k - y \rangle + (1 - \alpha_k) \langle (I - T)(x^k), x^k - y \rangle.$$

Upon letting  $k \rightarrow \infty$  in the previous inequality and utilizing the Assumption that  $\tilde{\delta} = \infty \iff \lim_k \frac{\alpha_k}{\beta_k} = 0$ , part (b) of Lemma 8 ( $y_k \rightarrow 0$ ), and also boundedness of

$\{x^k\}$ , we are led to the following

$$\limsup_k \langle (I - T)x^k, x^k - y \rangle \leq 0, \quad \forall y \in \text{Fix}(W),$$

which upon executing the limit,

$$\langle (I - T)x, x - y \rangle \leq 0, \quad \forall y \in \text{Fix}(W),$$

and which in turn yields  $x \in VI(T, \text{Fix}(W))$ .  $\square$

**Fact 10** If the interior of  $X^*$  is non-empty then  $X^* = \text{Fix}(T) \cap \text{Fix}(W)$ , and as well  $\{x^k\}$  is bounded.

**Proof** First, we show  $X^* = \text{Fix}(T) \cap \text{Fix}(W)$ . Let there exist  $x_0 \in \text{int}X^*$  and let  $x \in \mathbb{R}^n$  be given. Hence for sufficiently small  $t \in (0, 1)$ , we have that  $x_0 + t(x - x_0) \in X^* \subset \text{Fix}(W)$ , which further implies

$$\phi_1(x_0) = \phi_1(x_0 + t(x - x_0)) \leq (1 - t)\phi_1(x_0) + t\phi_1(x),$$

and so  $\phi_1(x_0) \leq \phi_1(x)$ . This means that  $x_0 \in \text{Fix}(T)$ . Therefore,  $\text{int}X^* \subseteq \text{Fix}(T)$ . On the other hand, since  $X^*$  is closed and convex, we therefore have

$$X^* = \text{cl}(\text{int}(X^*)) \subseteq \text{Fix}(T).$$

Consequently, as we already have  $X^* \cap \text{Fix}(T) = \text{Fix}(T) \cap \text{Fix}(W)$ , one can deduce that  $X^* = \text{Fix}(T) \cap \text{Fix}(W)$ , which verifies the desired equality. Notably, as mentioned in Remark 7, it is evident that the sequence  $x^k$  is bounded.  $\square$

**Assumption 11** (Quadratic growth condition) Suppose now that  $\phi_1$  grows quadratically (globally) away from a part of its minimizing set  $Y^* = \text{Fix}(W)$ , i.e.,  $X^*$ , meaning there is a real number  $\mu > 0$  such that

$$\phi_1(x) \geq \phi_1^* + \frac{\mu}{2} \text{dist}^2(x, X^*) \quad \forall x \in \Omega_1 \setminus \text{Fix}(W) \quad (17)$$

where  $\Omega_1 = \mathcal{B}(0, C_{x_0})$  for given  $x_0 \in \text{Fix}(W)$  and  $\phi_1^*$  represents the optimal value of  $\phi_1$ .

The quadratic growth condition can be interpreted as a notion of sharpness assumption on the function  $\phi_1$ , which describes functions that exhibit at least the behavior of  $\text{dist}(x, X^*)$ .

Originally introduced to establish the convergence of trajectories for the gradient flow of analytic functions, Bolte et al. proposed an extension to non-smooth functions in their work published in Bolte et al. (2007).

As a simple example, let us assume that  $\phi_1(x, y) = 0$  for  $(x, y) \in [-1, 1] \times [-1, 1]$  and for otherwise  $\phi_1(x, y) \geq 2$  and  $\Omega_1 = \mathcal{B}_2(0, 0)$ , we get  $\text{Fix}(W) = [-1, 1] \times [-1, 1]$  now, considering  $X^* = [-\frac{1}{2}, \frac{1}{2}] \times [-\frac{1}{2}, \frac{1}{2}]$ , we will observe, through a straightforward investigation, that (17) is verified.

**Theorem 12** *Let Assumption 11 hold, and  $\delta = 0$ . Then  $\{x^k\}$  converges to some  $x^* \in X^*$  such that*

$$\langle x^* - S(x^*), x - x^* \rangle \geq 0 \quad \forall x \in X^*.$$

**Proof** Strong convexity of  $\omega$ , and contractivity of the operator  $S$ , together implies that there is unique  $x^* \in X^*$  such that  $x^* = P_{X^*} Sx^*$  and  $x^* \in VI(S, X^*)$ , i.e.,

$$\langle x^* - S(x^*), x - x^* \rangle \geq 0, \quad \forall x \in X^*. \quad (18)$$

Since the sequence  $\{x^k\}$  is bounded, one sees that  $x^k \in \Omega_1$ . Furthermore, utilizing assumption 11, one may be readily verified that  $w(x^k) \subset X^*$ . Moreover, one can extract a convergent sub-sequence  $\{x^{k_i}\}$  of  $\{x^{k+1}\}$  or any sub-sequence thereof to  $x' \in X^*$ , which holds by Lemma 8, part c, and (69) so that

$$\begin{aligned} \limsup_k \langle S(x^*) - x^*, x^{k+1} - x^* \rangle &= \lim_i \langle S(x^*) - x^*, x^{k_i} - x^* \rangle \\ &= \langle S(x^*) - x^*, x' - x^* \rangle \leq 0. \end{aligned} \quad (19)$$

Next, we show  $x^k \rightarrow x^*$ . Let the sequences  $c_k$  and  $d_k$  are defined as

$$\begin{aligned} c_k &:= \alpha_k(S(x^k) - S(x^*)) + (1 - \alpha_k)\beta_k(T(x^k) - T(x^*)) \\ &\quad + (1 - \alpha_k)(1 - \beta_k)(W(x^k) - W(x^*)), \\ d_k &:= \alpha_k(S(x^*) - x^*) + (1 - \alpha_k)\beta_k(T(x^*) - x^*). \end{aligned}$$

From above it is immediate that  $c_k + d_k = x^{k+1} - x^*$ , and  $\|c_k\| \leq (1 - (1 - r)\alpha_k)\|x^k - x^*\|$ . By a simple calculation, one has

$$\|c_k + d_k\|^2 \leq \|c_k\|^2 + 2 \langle d_k, c_k + d_k \rangle,$$

and finally by plugging  $c_k$  and  $d_k$  in the previous inequality follows that

$$\begin{aligned} \|x^{k+1} - x^*\|^2 &\leq (1 - (1 - r)\alpha_k)\|x^k - x^*\|^2 + 2\alpha_k \langle S(x^*) - x^*, x^{k+1} - x^* \rangle \\ &\quad + 2(1 - \alpha_k)\beta_k \langle T(x^*) - x^*, x^{k+1} - x^* \rangle. \end{aligned} \quad (20)$$

Now, setting

$$\begin{cases} a_k = \|x^k - x^*\|^2, \\ \gamma_k = (1 - r)\alpha_k, \\ \delta_k = 2\alpha_k \langle S(x^*) - x^*, x^{k+1} - x^* \rangle + 2(1 - \alpha_k)\beta_k \langle T(x^*) - x^*, x^{k+1} - x^* \rangle. \end{cases} \quad (21)$$

One has that

$$a_{k+1} \leq (1 - \gamma_k)a_k + \delta_k.$$

Also, using the boundedness of  $\{x^k\}$  together with  $\delta = 0$  we can conclude that  $\limsup_k \frac{\delta_k}{\gamma_k} \leq 0$ . Indeed, taking into account (19) and  $\delta = 0$  gives

$$\begin{aligned} & \limsup_k \frac{\delta_k}{\gamma_k} \\ &= \limsup_k \left[ \frac{2\alpha_k \langle S(x^*) - x^*, x^{k+1} - x^* \rangle + 2(1 - \alpha_k)\beta_k \langle T(x^*) - x^*, x^{k+1} - x^* \rangle}{(1 - r)\alpha_k} \right] \\ &\leq \frac{2}{1 - r} \limsup_k \langle S(x^*) - x^*, x^{k+1} - x^* \rangle \\ &\quad + \frac{2}{1 - r} \limsup_k (1 - \alpha_k) \frac{\beta_k}{\alpha_k} \langle T(x^*) - x^*, x^{k+1} - x^* \rangle \leq 0. \end{aligned}$$

The desired assertion now follows from Lemma 5.  $\square$

**Theorem 13** *Let Assumption 11 hold, and  $\delta = \infty$ . Moreover, assume that  $\{x^k\}$  is bounded. Then  $\{x^k\}$  converges to some  $x^* \in X^*$  such that*

$$\langle x^* - S(x^*), x - x^* \rangle \geq 0 \quad \forall x \in X^*,$$

i.e.,  $\min_{x \in X^*} \omega(x) = \omega(x^*)$ .

**Proof** As before, there is a unique  $x^* \in X^*$  fixed point of the contraction map  $P_{X^*}S$ , i.e.,  $x^* = P_{X^*}Sx^*$ . Therefore  $x^* \in VI(S, X^*)$  and

$$\langle x^* - S(x^*), x - x^* \rangle \geq 0, \quad \forall x \in X^*$$

as a same method, due to the boundedness of  $x^k$  we may get a subsequence  $\{x^{k_i}\}$  converges to  $x' \in X^*$  such that

$$\begin{aligned} \limsup_k \langle S(x^*) - x^*, x^{k+1} - x^* \rangle &= \lim_i \langle S(x^*) - x^*, x^{k_i} - x^* \rangle \\ &= \langle S(x^*) - x^*, x' - x^* \rangle \leq 0, \end{aligned}$$

and also there is subsequence  $x^{k_j}$  converges to  $x'' \in X^*$  such that

$$\begin{aligned} \limsup_k \langle T(x^*) - x^*, x^{k+1} - x^* \rangle &= \lim_i \langle T(x^*) - x^*, x^{k_i} - x^* \rangle \\ &= \langle T(x^*) - x^*, x'' - x^* \rangle \leq 0. \end{aligned} \tag{22}$$

We show the last inequality. Using Lemma 9, one derives that  $x'' \in VI(I - T, \text{Fix}(W))$ , i.e.,

$$\langle T(x'') - x'', x - x'' \rangle \leq 0 \quad \forall x \in \text{Fix}(W)$$

Taking  $x = x^* \in X^* \subseteq \text{Fix}(W)$ , gives  $\langle T(x'') - x'', x^* - x'' \rangle \leq 0$ . Using monotonicity of  $I - T$  yields that  $\langle T(x^*) - x^*, x'' - x^* \rangle \leq 0$  and this follows (22). The rest of the proof follows from (20) and (21), and Lemma 5.  $\square$

As another application of Theorem 12, one may point to Theorem 6.1 of Xu (2002) for solving the following quadratic minimization problem:

$$\min_{x \in K} [\omega(x) := \frac{\mu}{2} \langle Ax, x \rangle + \frac{1}{2} \|x - u\|^2 - \langle x, b \rangle], \quad (23)$$

where  $K$  is a nonempty closed convex and  $\mu \geq 0$  is a real number,  $u, b \in \mathbb{R}^n$  and  $A$  is a bounded linear operator which is positive ( $\langle Ax, x \rangle \geq 0$  for all  $x \in \mathbb{R}^n$ ). Set  $S(x) := x - r \nabla \omega(x)$  and  $T(x) := \text{prox}_{\delta K}(x) = P_K(x)$ . Then the sequence  $\{x^k\}$  generated by  $x^{k+1} = \alpha_k S(x^k) + (1 - \alpha_k) T(x^k)$  converges to the unique solution  $x^*$  of problem (23) under the mild assumption  $\alpha_k \rightarrow 0$  and  $\sum_k \alpha_k = \infty$ . We drop

the assumption that  $\lim_k \frac{\alpha_{k+1}}{\alpha_k} = 1$ . Notice that when we take  $K = \mathbb{R}^n$ , then problem (23) reduces to a classical convex quadratic optimization problem, in which case  $x^k \rightarrow \text{prox}_f(u)$  where  $f(u) = \frac{1}{2} \langle Au, u \rangle + \langle u, b \rangle$  and  $\text{prox}_f(u) = (A + I)^{-1}(u - b)$ .

**Remark 14** Knowing relation (20) and Assumption 2, we find out that the two following conditions together imply the convergence of the sequence  $\{x^k\}$ :

$$\limsup_k \langle (S - I)x^*, x^k - x^* \rangle \leq 0, \quad (24)$$

and

$$\limsup_k \frac{\beta_k}{\alpha_k} \langle (T - I)x^*, x^k - x^* \rangle \leq 0. \quad (25)$$

Thanks to the Assumptions of Theorem 12,  $x^*$  solves  $VI(S, X^*)$ . This is due to the fact that  $\delta = 0$ , which means  $\beta_k \rightarrow 0$  faster than  $\alpha_k \rightarrow 0$ . Afterwards, the term  $\alpha_k S(x^k)$  dominates, while the term  $\beta_k T(x^k)$  becomes negligible. When  $\delta = \infty$ , it is difficult to confirm the verification of condition (25) without assuming bounded linear regularity to control the growth of  $\|x - T(x)\|$ .

Up to now, we have shown that the sequence  $\{x^k\}$  is bounded and convergent, provided that  $\delta = 0$ . A natural question is to ask whether the sequence  $\{x^k\}$  is convergent when  $\delta$  is non-zero. The following proposition guarantees, under the assumption  $\delta \in [0, +\infty)$ , that there is a particular variational inequality that is satisfied for any limit point of the sequence generated by the Algorithm.



**Proposition 15** Assume  $\delta < \infty$ , together with Assumptions 2, 3 and 4. Then sequence  $\{x^k\}$  converges to the unique solution of the variational inequality

$$\exists \tilde{x} \in \text{Fix}(W) \quad \langle (I - S)\tilde{x} + \delta(I - T)\tilde{x}, x - \tilde{x} \rangle \geq 0, \quad \forall x \in \text{Fix}(W). \quad (26)$$

**Proof** Set  $y_k = \frac{1}{\beta_k}(x^k - x^{k+1})$ . From part (b) of Lemma 8 we have  $y_k \rightarrow 0$  as  $k \rightarrow \infty$ . By the definition of iteration (12) and monotonicity of  $I - W$  for all  $y \in \text{Fix}(W)$ , one sees easily that

$$\begin{aligned} \langle y^k, x^k - y \rangle &= \frac{\alpha_k}{\beta_k} \langle (I - S)x^k, x^k - y \rangle + (1 - \alpha_k) \langle (I - T)x^k, x^k - y \rangle \\ &\quad + \frac{(1 - \alpha_k)(1 - \beta_k)}{\beta_k} \langle (I - W)x^k - (I - W)y, x^k - y \rangle \\ &\geq \frac{\alpha_k}{\beta_k} \langle (I - S)x^k, x^k - y \rangle + (1 - \alpha_k) \langle (I - T)x^k, x^k - y \rangle, \end{aligned} \quad (27)$$

which implies that

$$\frac{\beta_k}{\alpha_k} \langle y^k, x^k - y \rangle \geq \langle (I - S)x^k, x^k - y \rangle + \frac{\beta_k(1 - \alpha_k)}{\alpha_k} \langle (I - T)x^k, x^k - y \rangle. \quad (28)$$

Now, for given  $x_1, x_2 \in w(\{x^k\})$ , there exist sub-sequences  $\{x^{k_i}\}$  and  $\{x^{k_j}\}$  of  $\{x^k\}$ , such that  $x^{k_i} \rightarrow x_1$  and  $x^{k_j} \rightarrow x_2$ . On taking the limsup of (28) and using the fact that  $y_k \rightarrow 0$  and  $\limsup_k \frac{\beta_k}{\alpha_k} = \delta \in [0, \infty)$ , we deduce that,

$$\begin{cases} \langle \delta(I - T)x_1 + (I - S)x_1, x_1 - y \rangle \leq 0, \\ \langle \delta(I - T)x_2 + (I - S)x_2, x_2 - y \rangle \leq 0, \end{cases} \quad \forall y \in \text{Fix}(W). \quad (29)$$

Rearranging (29) by substituting  $y = x_1$  and  $y = x_2$  shows that

$$\langle (I - S)x_1, x_1 - x_2 \rangle \leq -\delta \langle (I - T)x_1, x_1 - x_2 \rangle, \quad (30)$$

$$-\langle (I - S)x_2, x_1 - x_2 \rangle \leq \delta \langle (I - T)x_2, x_1 - x_2 \rangle. \quad (31)$$

On the other hand, since  $I - S$  is  $(1 - r)$ -strongly monotone and  $I - T$  is monotone by adding up inequalities (30) and (31), one obtains that

$$\begin{aligned} (1 - r)\|x_1 - x_2\|^2 &\leq \langle (I - S)x_1 - (I - S)x_2, x_1 - x_2 \rangle \\ &\leq -\delta \langle (I - T)x_1 - (I - T)x_2, x_1 - x_2 \rangle \\ &\leq 0. \end{aligned}$$

So,  $x_1 = x_2$ . This shows that  $\{x^k\}$  converges. (Here, we have used the fact that the sequence  $\{x^k\}$  converges if and only if every sub-sequence of  $\{x^k\}$  contains a convergent sub-sequence.) Setting  $\tilde{x} := \lim_{k \rightarrow \infty} x^k$ , we then see from (29) that

$$\langle (I - S)\tilde{x} + \delta(I - T)\tilde{x}, x - \tilde{x} \rangle \geq 0, \quad \forall x \in \text{Fix}(W).$$

This completes the proof.  $\square$

**Corollary 16** For each operator  $P \in [I - S, I - T]$  one has  $VI(P, \text{Fix}(W)) \neq \emptyset$  where  $c \in [a, b]$  means that there is  $t \in [0, 1]$  such that  $c = ta + (1 - t)b$ .

## 2.4 Convergence analysis under an error-bound condition

Here, we introduce an error-bound condition that facilitates additional convergence guarantees. Let us denote the closed ball of radius  $\rho$  centred at 0 by  $\mathcal{B}(0, \rho)$ .

**Definition 17** (*Error bound condition*), Borwein et al. (2017) Let  $W : X \rightarrow X$  be such that  $\text{Fix}(W) \neq \emptyset$ . We say that  $W$  is boundedly linearly regular if

$$\forall \rho > 0 \ (\exists \theta > 0) \ (\forall x \in \mathcal{B}(0; \rho)) \ d(x, \text{Fix}(W)) \leq \theta \|x - W(x)\|.$$

note that in general  $\theta$  depends on  $\rho$ , which we sometimes indicate by writing  $\theta = \theta(\rho)$ .

The notion of a bounded linear regularity is a valuable property in optimization and variational analysis. It ensures that a function behaves well near its critical points, and has been used in Bauschke et al. (2015) to analyze linear convergence of algorithms involving nonexpansive mappings. An exemplary and practically significant illustration of an objective that is non-quasi-strongly convex yet satisfies the quadratic growth condition is the LASSO problem:

$$\min_{x \in \mathbb{R}^n} \left[ \frac{1}{2} \|Qx - b\|^2 + \lambda \|x\|_1 \right] \quad (32)$$

when the operator  $Q$  has a nontrivial kernel. Further classes of functions that possess a regular error-bound property include the following:

### Example 2

$$\begin{aligned} X = \mathbb{R}^n, \quad \|\cdot\| = \|\cdot\|_2 \quad f(x) &= \|Ax - b\|_2^2, \quad g(x) = \|x\|_1 \\ X = \mathbb{R}^n, \quad \|\cdot\| = \|\cdot\|_2 \quad f(x) &= \|Ax - b\|_2^2, \quad g(x) = \begin{cases} 0 & x \in \mathcal{B}_2^{\frac{n}{2}} \\ +\infty & \text{else} \end{cases} \\ \|\cdot\| = \|\cdot\|_F, \quad X = \mathbb{R}^{p \times n}, \quad f(x) &= \|Ax - b\|_2^2, \quad g(x) = \|x\|_{\text{nuc}} \end{aligned}$$

**Proposition 18** For all  $t \in \left(0, \frac{1}{L_{f_2}}\right]$  and  $x \in \text{dom}(\partial\phi_2)$  one has

$$\|x - T_t(x)\| \leq td(0, \partial\phi_2(x)).$$

**Proof** For every  $x \in \text{dom}(\partial\phi_2)$  we have  $\partial\phi_2(x) = \nabla f_2(x) + \partial g_2(x)$ . For given  $t \in (0, \frac{1}{L_{f_2}}]$  and  $z \in \partial\phi_2(x)$  one has

$$(x - t\nabla f_2(x)) + tz \in x + t\partial g_2(x) = (I + t\partial g_2)(x),$$

we must then have

$$(I + t\partial g_2)^{-1}((x - t\nabla f_2(x)) + tz) = x,$$

or equivalently,  $T_t(x + tz) = \text{prox}_{tg_2}(x + tz - t\nabla f_2(x)) = x$ . Since the proximal mapping is non-expansive, we deduce that

$$\|x - T_t(x)\| = \|T_t(x + tz) - T_t(x)\| \leq t\|z\|, \quad \forall z \in \partial\varphi_2(x).$$

Letting  $z$  be the minimal norm element of  $\partial\varphi_2(x)$ , we derived the claimed inequality  $\|x - T_t(x)\| \leq td(0, \partial\varphi_1(x))$ .  $\square$

We shall now study cases wherein  $\delta$  is not finite. From now on, we use  $\Omega := VI(T, \text{Fix}(W))$  and assume it is non-empty.

**Theorem 19** Assume  $\tilde{\delta} = \infty$ , together with Assumptions 2, 3, and 4. Assume also that  $\{x^k\}$  is bounded. Moreover, if the following assumptions hold

(A<sub>2</sub>)  $W$  is boundedly linearly regular,

(A<sub>3</sub>)  $\limsup_k \frac{\beta_k^2}{\alpha_k} = 0$ ,

then the sequence  $\{x^k\}$  converges to  $x^*$ , the unique solution of

$$\langle x^* - S(x^*), x - x^* \rangle \geq 0, \quad \forall x \in \Omega. \quad (33)$$

Furthermore, this implies that  $x^*$  minimizes  $\omega$  over  $\Omega$ , i.e.,

$$\min_{x \in \Omega} \omega(x) = \omega(x^*).$$

**Proof** Since  $\Omega$  is closed and convex and  $S$  is a contraction, there exists  $x^* \in \Omega$ , which is a unique fixed point of the projection map  $P_\Omega S(x^*) = x^*$ , i.e.,

$$\langle x^* - S(x^*), x - x^* \rangle \geq 0, \quad \forall x \in \Omega. \quad (34)$$

To deduce  $x^k \rightarrow x^*$ , we first note that  $x^* \in \Omega := VI(T, \text{Fix}(W))$ , which implies that

$$\langle (T - I)x^*, x - x^* \rangle \leq 0 \quad \forall x \in \text{Fix}(W),$$

and since  $P_{\text{Fix}(W)}(x^k) \in \text{Fix}(W)$ , one gets

$$\langle (T - I)x^*, P_{\text{Fix}(W)}(x^{k+1}) - x^* \rangle \leq 0. \quad (35)$$

On the other hand, since  $\{x^k\}$  is bounded, there exists  $M > 0$  and  $k_0 \geq 0$  such that for all  $k \geq k_0$ , one has  $x^k \in \mathcal{B}(0, M)$ . So, by applying Assumption (A<sub>1</sub>), one can easily

observe that there exists  $\theta > 0$  such that

$$d(x^k, \text{Fix}(W)) \leq \theta \|x^k - W(x^k)\|. \quad (36)$$

Employing (35) and (36), one has

$$\begin{aligned} & \left\langle (T - I)x^*, x^{k+1} - x^* \right\rangle \\ &= \left\langle (T - I)x^*, x^{k+1} - P_{\text{Fix}(W)}(x^{k+1}) + P_{\text{Fix}(W)}(x^{k+1}) - x^* \right\rangle \\ &= \left\langle (T - I)x^*, x^{k+1} - P_{\text{Fix}(W)}(x^{k+1}) \right\rangle \\ &\quad + \left\langle (T - I)x^*, P_{\text{Fix}(W)}(x^{k+1}) - x^* \right\rangle \\ &\leq \left\langle (T - I)x^*, x^{k+1} - P_{\text{Fix}(W)}(x^{k+1}) \right\rangle \\ &\leq \|T(x^*) - x^*\| \|x^{k+1} - P_{\text{Fix}(W)}(x^{k+1})\| \\ &= \|T(x^*) - x^*\| d(x^{k+1}, \text{Fix}(W)) \\ &\leq \theta \|T(x^*) - x^*\| \|x^{k+1} - W(x^{k+1})\|. \end{aligned}$$

Hence

$$\left\langle (T - I)x^*, x^{k+1} - x^* \right\rangle \leq \theta \|T(x^*) - x^*\| \|x^{k+1} - W(x^{k+1})\|. \quad (37)$$

Now, since  $\{x^k\}$  is bounded, one can find a constant  $C > 0$  so that

$$C \geq \sup_{k \geq 1} \left\{ \|S(x^k)\|, \|T(x^k)\|, \|W(x^k)\| \right\},$$

and we will then have

$$\begin{aligned} & \|x^{k+1} - Wx^{k+1}\| \\ &\leq \|x^{k+1} - W(x^k)\| + \|W(x^{k+1}) - W(x^k)\| \\ &\leq \|x^{k+1} - Wx^k\| + \|x^{k+1} - x^k\| \\ &\leq \alpha_k \|S(x^k)\| + \beta_k \|T(x^k)\| + (\alpha_k + \beta_k + \alpha_k \beta_k) \|W(x^k)\| + \|x^{k+1} - x^k\| \\ &\leq (2\alpha_k + 2\beta_k + \alpha_k \beta_k)C + \|x^{k+1} - x^k\|. \end{aligned}$$

Therefore, by combining the previous inequality and (37), we get:

$$\left\langle (T - I)x^*, x^{k+1} - x^* \right\rangle \leq \theta \|T(x^*) - x^*\| \left[ (2\alpha_k + 2\beta_k + \alpha_k \beta_k)C + \|x^{k+1} - x^k\| \right]. \quad (38)$$

Now, multiplication (38) with  $\frac{\beta_k}{\alpha_k}$  yields

$$\begin{aligned} & \frac{\beta_k}{\alpha_k} \left\langle (T - I)x^*, x^{k+1} - x^* \right\rangle \\ & \leq \theta \|T(x^*) - x^*\| \left[ \left( 2\beta_k + 2\frac{\beta_k^2}{\alpha_k} + \beta_k^2 \right) C + \frac{\beta_k^2}{\alpha_k} \cdot \frac{\|x^{k+1} - x^k\|}{\beta_k} \right]. \end{aligned} \quad (39)$$

Using (39), Assumption (A<sub>4</sub>), and part (b) of Lemma 8, we will observe that

$$\limsup_k \frac{\beta_k}{\alpha_k} \left\langle (T - I)x^*, x^{k+1} - x^* \right\rangle \leq 0. \quad (40)$$

Moreover by Lemma 9, we have  $w(\{x^k\}) \subseteq \Omega$ . Now, since  $\{x^{k+1}\}$  is bounded there exists a convergent sub-sequence  $\{x^{k_i}\}$  of  $\{x^{k+1}\}$  to  $x' \in \Omega$ . From (34), it can be seen that

$$\begin{aligned} \limsup_k \left\langle (S - I)x^*, x^{k+1} - x^* \right\rangle &= \lim_k \left\langle (S - I)x^*, x^{k_i} - x^* \right\rangle \\ &= \left\langle (S - I)x^*, x' - x^* \right\rangle \leq 0. \end{aligned} \quad (41)$$

Recall that we still have inequality (20). By a similar argument as in Remark 14, from (40) and (41) and in view of Lemma 5, we see that  $x^k \rightarrow x^*$  and the proof is complete.  $\square$

Notice that by Lemma 9, we know that when  $\tilde{\delta} = \infty$ , then  $w(\{x^k\}) \subseteq \Omega$ . The following example shows that this is not a necessary condition.

**Example 3** Take the choices

$$S(x) = \frac{x}{4}, \quad T(x) = \begin{cases} x & x \in [-1, 1], \\ 1 & x \geq 1, \\ -1 & x \leq -1, \end{cases}, \quad W(x) = \begin{cases} x & x \in [0, 2], \\ 2 & x \geq 2, \\ 0 & x \leq 0. \end{cases}$$

Also, consider  $\alpha_k = \frac{1}{k}$  and  $\beta_k = \frac{1}{k^2}$ . It can be seen that  $\delta = 0$  and  $\Omega = VI(T, \text{Fix}(W)) = [0, 1]$  and also  $w(\{x^k\}) = \{0\} \subset \Omega$ .

The following result asserts the existence of a limit point satisfying a variational inequality under a mild assumption related to the preceding theorem without any condition on  $\delta$ .

**Proposition 20** Assume  $\delta < \infty$ , together with Assumptions 2, 3, 4, and (A<sub>2</sub>) of Theorem 19. Moreover, assume that Assumption 11 holds for  $\Omega$  in replace of  $X^*$ . Then the sequence  $\{x^k\}$  converges to  $x^*$ , which is the unique solution of

$$\langle x^* - S(x^*), x - x^* \rangle \geq 0, \quad \forall x \in \Omega. \quad (42)$$

**Proof** This is immediate from Theorem (19).  $\square$

The following fact provides the limit of distance between  $\{x^k\}$  and  $\text{Fix}(W)$  and  $X^*$ , respectively.

**Fact 21** Let Assumption 4 hold and  $h_k = d(x^k, X^*)$ . Also suppose that  $\{x^k\}$  is bounded. Then the following assertion holds

- a) If  $h_k \neq 0$  then  $\lim_{k \rightarrow \infty} \frac{(h_{k+1}^2 - h_k^2)^+}{h_{k+1}} = 0$ .  
 b)  $\lim_{k \rightarrow \infty} d(x^k, \text{Fix}(W)) = 0$ .

**Proof** We just prove the first assertion. (The second is straightforward from the boundedness of  $\{x^k\}$ .) The proof relies on the study of the sequence  $\{h_k\}$ . Since  $P_{X^*}$  is the projection operator onto the convex set  $X^*$ , we have

$$\begin{aligned} \frac{1}{2}h_k^2 &= \frac{1}{2}\|x^k - P_{X^*}(x^k)\|^2 \\ &= \frac{1}{2}\|(x^k - P_{X^*}(x^k)) - (x^{k+1} - P_{X^*}(x^{k+1})) + (x^{k+1} - P_{X^*}(x^{k+1}))\|^2 \\ &= \frac{1}{2}\|(x^k - P_{X^*}(x^k)) - (x^{k+1} - P_{X^*}(x^{k+1}))\|^2 + \frac{1}{2}\|x^{k+1} - P_{X^*}(x^{k+1})\|^2 \\ &\quad + \left\langle (x^k - P_{X^*}(x^k)) - (x^{k+1} - P_{X^*}(x^{k+1})), x^{k+1} - P_{X^*}(x^{k+1}) \right\rangle \\ &\geq \frac{1}{2}h_{k+1}^2 + \left\langle x^k - x^{k+1}, x^{k+1} - P_{X^*}(x^{k+1}) \right\rangle \\ &\quad + \left\langle P_{X^*}(x^{k+1}) - P_{X^*}(x^k), x^{k+1} - P_{X^*}(x^{k+1}) \right\rangle. \end{aligned}$$

Now,  $P_{X^*}(x^k) \in X^*$  and consequently

$$\left\langle P_{X^*}(x^{k+1}) - P_{X^*}(x^k), x^{k+1} - P_{X^*}(x^{k+1}) \right\rangle \geq 0.$$

Therefore

$$\begin{aligned} \frac{1}{2}h_{k+1}^2 - \frac{1}{2}h_k^2 &\leq \left\langle x^{k+1} - x^k, x^{k+1} - P_{X^*}(x^{k+1}) \right\rangle \\ &\leq \left\langle x^{k+1} - x^k, x^{k+1} - P_{X^*}(x^{k+1}) \right\rangle \\ &\leq \|x^{k+1} - x^k\| \|x^{k+1} - P_{X^*}(x^{k+1})\| \\ &= \|x^{k+1} - x^k\| h_{k+1}. \end{aligned} \tag{43}$$

Since  $h_{k+1} \neq 0$  the last inequality follows that

$$0 \leq \frac{(h_{k+1}^2 - h_k^2)^+}{h_{k+1}} \leq 2 \|x^k - x^{k+1}\|.$$

Finally, the proof is completed by part (a) of Lemma 8.  $\square$

**Remark 22** We would also like to point out that if in Theorems 19 and Proposition 20 we had  $\Omega := VI(T, \text{Fix}(W)) = X^*$  then the trilevel optimization problem (2) would have a solution.

**Remark 23** Our assumptions on  $\alpha_k$  and  $\beta_k$  are weaker than of assumptions (Maingé and Abdellatif 2007; Moudafi 2007). For instance, consider that  $\lim_k \frac{\beta_k}{\alpha_k}$  need not exist.

Instead, we consider  $\limsup_k \frac{\beta_k}{\alpha_k}$ . See Example 4 below, where there is no limit  $\frac{\beta_k}{\alpha_k}$ , but our results still apply.

**Example 4** Let  $\beta_k = \frac{1}{(k+1)(3+(-1)^k)}$ ,  $\alpha_k = \frac{1}{(k+1)(2+(-1)^k)}$  clearly  $\lim_k \frac{\beta_k}{\alpha_k}$  does not exist, but  $\limsup_k \frac{\beta_k}{\alpha_k} = \frac{3}{4}$  and also

$$\sum_{k=1}^{\infty} \alpha_k = \infty, \quad \alpha_k \rightarrow 0.$$

Now, we are ready to give an example related to the step-sizes  $\alpha_k$  and  $\beta_k$  that guarantee the convergence  $\{x^k\}$  of all of our results. Note that in all cases  $\lim_k \frac{\beta_k}{\alpha_k}$  may not exist.

**Example 5** Consider  $\alpha_k = \frac{1}{(k+1)^\lambda(2+(-1)^k)}$  and  $\beta_k = \frac{1}{(k+1)^\gamma(3+(-1)^k)}$  with  $\lambda > 0$ ,  $\gamma > 0$ . Now for large sufficient  $k$ , we have the following estimation

$$|\alpha_{k+1} - \alpha_k| \approx \frac{1}{(k+1)^\lambda}, \quad |\beta_{k+1} - \beta_k| \approx \frac{1}{(k+1)^\gamma}.$$

It is easy to check that

$$\delta = \limsup_k \frac{\beta_k}{\alpha_k} = \begin{cases} 0 & \lambda < \gamma, \\ 1 & \lambda = \gamma, \\ +\infty & \lambda > \gamma. \end{cases}$$

Furthermore,

- Assumption 2 holds when  $0 < \lambda \leq 1$ .
- Assumption 3 holds when  $0 < \lambda + \gamma \leq 1$ .
- Assumption 4 holds when  $0 < \lambda, \gamma < 1$ .
- Assumption (A<sub>4</sub>) holds when  $\lambda \leq 2\gamma$ .

Remark that for the case  $\tilde{\delta} = \infty$ , it is sufficient to consider  $\alpha_k = \frac{1}{(k+1)^\lambda}$  and  $\beta_k = \frac{1}{(k+1)^\gamma}$  with  $\lim_k \frac{\beta_k}{\alpha_k} = \infty$  for  $\lambda > \gamma$ . Now we are ready to present a taxonomy of assumptions with respect to  $\alpha_k$  and  $\beta_k$ , as referenced in Table 2

- (a):  $\lambda < \gamma$ .

- (b):  $0 < \lambda \leq \gamma < 1$  or  $0 < \gamma < \lambda < 1$ .
- (c):  $0 < \lambda \leq \gamma < 1$  and  $0 < \lambda + \gamma < 1$ .
- (d):  $0 < \gamma < \lambda \leq 1$ ,  $0 < \lambda + \gamma \leq 1$  and  $\lambda \leq 2\gamma$ .
- (e):  $0 < \gamma < \lambda \leq 1$ ,  $0 < \lambda + \gamma \leq 1$ .
- (f):  $0 < \gamma = \lambda < 1$ .

### 3 Convergence rate analysis for trilevel optimization problems

In this section, we present the main result of this paper. This addresses the rate of convergence of the sequence  $\{x^k\}$  generated by Algorithm 1 with a particular choice of step-sizes.

#### 3.1 Technical lemmas

The technical lemma which we state next, and for which we refer to Sabach and Shtern (2017), will play a crucial role in the convergence analysis.

**Lemma 24** (Sabach and Shtern 2017, Lemma 3) *Let  $M > 0$ . Suppose that  $\{a_k\}$  is a sequence of non-negative real numbers which satisfy  $a_1 \leq M$  and*

$$a_{k+1} \leq (1 - \gamma b_{k+1})a_k + (b_k - b_{k+1})c_k \quad k \geq 1,$$

where  $\gamma \in (0, 1]$ ,  $\{b_k\}$  is a sequence defined as  $b_k := \min\{\frac{2}{\gamma k}, 1\}$  and  $\{c_k\}$  is a sequence of real numbers such that  $c_k \leq M < \infty$ .

Then, the sequence  $\{a_k\}$  satisfies

$$a_k \leq \frac{MJ}{\gamma k} \quad k \geq 1, \quad \text{where } J = \left\lfloor \frac{2}{\gamma} \right\rfloor.$$

The next result will be useful for the rate of convergence.

**Lemma 25** *One has the following*

$$\phi_i(x - t\psi_i(x)) - \phi_i(y) \leq \frac{1}{2t}\|x - y\|^2, \quad \forall x, y \in \mathbb{R}^n, \quad t \in \left(0, \frac{1}{L_{f_i}}\right] \quad (44)$$

where  $\phi_i(x) := f_i(x) + g_i(x)$  and  $\psi_i(x) := \frac{1}{t}(x - \text{prox}_{t g_i}(x - t \nabla f_i(x)))$  and  $i = 1, 2$

**Proof** Let  $i = 1$ . Using Lipschitz continuity of  $f_1$  with parameter  $L_{f_1}$  it is well-known that convexity of  $f_1$  is equivalent to

$$f_1(y) \leq f_1(x) + \nabla f_1(x)^T(y - x) + \frac{L_{f_1}}{2}\|x - y\|^2, \quad \forall x, y \in \mathbb{R}^n. \quad (45)$$



Assume that  $x \in \mathbb{R}^n$  and  $t \in (0, \frac{1}{L_{f_1}}]$  be given. Plugging  $y = x - t\psi_1(x)$  in the previous inequality one obtains that

$$\begin{aligned} f_1(x - t\psi_1(x)) &\leq f_1(x) + \nabla f_1(x)^T((x - t\psi_1(x)) - x) + \frac{L_{f_1}t^2}{2}\|\psi_1(x)\|^2 \\ &\leq f_1(x) - t\nabla f_1(x)^T(\psi_1(x)) + \frac{t}{2}\|\psi_1(x)\|^2. \end{aligned}$$

Now from  $x - t\psi_1(x) = \text{prox}_{tg_1}(x - t\nabla f_1(x))$ , we get to

$$\psi_1(x) - \nabla f_1(x) \in \partial g_1(x - t\psi_1(x)).$$

Therefore,

$$g_1(y) - g_1(x - t\psi_1(x)) \geq (\psi_1(x) - \nabla f_1(x))^T(y - x + t\psi_1(x)). \quad (46)$$

Now by simplifying and taking into account (46) one has

$$\begin{aligned} \phi_1(x - t\psi_1(x)) &= f_1(x - t\psi_1(x)) + g_1(x - t\psi_1(x)) \\ &\leq f_1(x) - t\nabla f_1(x)^T(\psi_1(x)) + \frac{t}{2}\|\psi_1(x)\|^2 + g_1(x - t\psi_1(x)) \\ &\leq f_1(y) - \nabla f_1(x)^T(y - x) - t\nabla f_1(x)^T(\psi_1(x)) + \frac{t}{2}\|\psi_1(x)\|^2 + g_1(x - t\psi_1(x)) \\ &\leq f_1(y) - \nabla f_1(x)^T(y - x) - t\nabla f_1(x)^T(\psi_1(x)) + \frac{t}{2}\|\psi_1(x)\|^2 \\ &\quad + g_1(y) - (\psi_1(x) - \nabla f_1(x))^T(y - x + t\psi_1(x)) \\ &\leq \phi_1(y) - \|\psi_1(x)\|\|x - y\| - t\|\psi_1(x)\|^2 \\ &\leq \phi_1(y) + \frac{1}{2t}[\|x - y\|^2 - \|(x - y) - t\psi_1(x)\|^2] \leq \phi_1(y) + \frac{1}{2t}\|x - y\|^2, \quad (47) \end{aligned}$$

and therefore it follows

$$\phi_1(x - t\psi_1(x)) - \phi_1(y) \leq \frac{1}{2t}\|x - y\|^2, \quad \forall x, y \in \mathbb{R}^n, \quad t \in \left(0, \frac{1}{L_{f_1}}\right].$$

□

We are now in a position to derive the following result which appeared in a similar form Beck and Teboulle (2009), Sabach and Shtern (2017), however for our context the proof had to be modified.

**Proposition 26** (Sabach and Shtern 2017, Proposition1) *Let  $x \in \mathbb{R}^n$  and denote  $x^+ = T_t(x)$ . Then*

$$\phi_2(x^+) - \phi_2(u) \leq \frac{1}{t}\langle x - x^+, x - u \rangle - \frac{1}{2t}\|x - x^+\|^2, \quad \forall (u, t) \in \mathbb{R}^n \times \left(0, \frac{1}{L_{f_2}}\right].$$

and also if  $z = W_s(x)$ . Then

$$\phi_1(z) - \phi_1(u) \leq \frac{2}{s} \langle x - z, x - u \rangle - \frac{1}{2s} \|x - z\|^2, \quad \forall (u, s) \in \mathbb{R}^n \times \left(0, \frac{1}{L_{f_1}}\right].$$

**Proof** We will just prove the first part. The second part can be proved by the same method. Assume that  $x^+ = T_t(x)$  so we have

$$\psi_2(x) = \frac{1}{t}(x - \text{prox}_{tg_2}(x - t\nabla f_2(x))) = \frac{1}{t}(x - T_t(x)) = \frac{1}{t}(x - x^+).$$

From (47) of Lemma 25 we obtain

$$\begin{aligned} & \phi_2(x - t\psi_1(x)) - \phi_2(u) \\ & \leq -\|\psi_1(x)\| \|x - u\| - t\|\psi_1(x)\|^2 \leq \langle \psi_1(x), x - u \rangle - \frac{t}{2}\|\psi_1(x)\|^2, \end{aligned}$$

and then by plugging  $\psi_1(x) = \frac{x-x^+}{t}$  in the previous inequality one get

$$\phi_2(x^+) - \phi_2(u) \leq \frac{1}{t} \langle x - x^+, x - u \rangle - \frac{1}{2t} \|x - x^+\|^2, \quad \forall (u, t) \in \mathbb{R}^n \times \left(0, \frac{1}{L_{f_2}}\right].$$

and the desired result follows.  $\square$

Now, we set up the sequences  $\alpha_k$  and  $\beta_k$  and define the constant  $J$  as

$$\alpha_k = \min \left\{ \frac{2}{(1-r)k}, 1 \right\}, \quad \beta_k = \frac{\alpha_k - \alpha_{k+1}}{2(2 - \alpha_k)}, \quad J = \left\lfloor \frac{2}{(1-r)} \right\rfloor, \quad k \geq 1 \quad (48)$$

where  $r \in (0, 1]$ . Clearly, Assumptions 2, 3, 4 are satisfied under (48).

To begin, we present the following lemma, which plays a key role in the sequel.

**Lemma 27** Assume that  $\{x^k\}$ ,  $\{y^k\}$ ,  $\{z^k\}$  and  $\{v^k\}$  be the sequences generated by Algorithm 1 and also  $x \in \text{Fix}(W)$  be given, defining  $y = T(x)$  and  $v = S(x)$ . Then, for every  $k \geq 1$  the following relations hold true.

$$\begin{cases} \|y^{k+1} - y^k\| \leq \|x^k - x^{k-1}\|, \\ \|z^{k+1} - z^k\| \leq \|x^k - x^{k-1}\|, \\ \|v^{k+1} - v^k\| \leq r\|x^k - x^{k-1}\|, \\ \|z^{k+1} - x\| \leq \|x^k - x\|, \\ \|y^{k+1} - y\| \leq \|x^k - x\|, \\ \|v^{k+1} - v\| \leq r\|x^k - x\|, \end{cases} \quad (49)$$

and there exists positive constants  $C_S, C_T$  and  $C_x$  so that

$$\begin{cases} \|y^k - z^k\| \leq C_T + 2C_x, \\ \|y^k - v^k\| \leq C_S + C_T + 2C_x, \\ \|v^k - z^k\| \leq C_S + 2C_x. \end{cases} \quad (50)$$

**Proof** All parts are a direct consequence of non-expansively of  $T$ ,  $W$  and the contraction property of  $S$  and Lemma 6.  $\square$

**Lemma 28** Let  $\{x^k\}$ ,  $\{y^k\}$ ,  $\{z^k\}$  and  $\{v^k\}$  be sequences generated by the Algorithm (1), where  $\{\alpha_k\}$  and  $\{\beta_k\}$  are defined by (48). Then for every  $x \in \text{Fix}(W)$  one has

$$\begin{cases} \|x^k - x^{k-1}\| \leq \frac{(C_S + 2C_T + 5C_x)J}{(1-r)k} \\ \text{and} \\ \|z^k - x^{k-1}\| \leq \frac{(C_S + 2C_T + 5C_x)(J+2)}{(1-r)k}, \end{cases} \quad (51)$$

where  $C_S$ ,  $C_T$ ,  $C_x$  are defined in Lemma 6 and  $J = \left\lfloor \frac{2}{(1-r)} \right\rfloor$ .

**Proof** One can write

$$\begin{aligned} x^{k+1} - x^k &= \alpha_{k+1}[v^{k+1} - v^k] \\ &\quad + (1 - \alpha_{k+1})[\beta_{k+1}(y^{k+1} - y^k) + (1 - \beta_{k+1})(z^{k+1} - z^k)] \\ &\quad + (\alpha_{k+1} - \alpha_k)[v^k - \beta_{k+1}y^k - (1 - \beta_k)z^k] \\ &\quad + (1 - \alpha_{k+1})(\beta_{k+1} - \beta_k)[y^k - z^k]. \end{aligned}$$

Now, one gets that

$$\begin{aligned} \|x^{k+1} - x^k\| &\leq (1 - (1-r)\alpha_{k+1})\|x^k - x^{k-1}\| \\ &\quad + (\alpha_k - \alpha_{k+1})\|v^k - \beta_{k+1}y^k - (1 - \beta_k)z^k\| \\ &\quad + (1 - \alpha_{k+1})(\beta_{k+1} - \beta_k)(\|y^k - z^k\|) \\ &\leq (1 - (1-r)\alpha_{k+1})\|x^k - x^{k-1}\| + (\alpha_k - \alpha_{k+1})c_k \end{aligned}$$

as well as one can easily follow that

$$\begin{aligned} c_k &:= \|v^k - \beta_{k+1}y^k - (1 - \beta_k)z^k + (1 - \alpha_{k+1})(\beta_{k+1} - \beta_k)(y^k - z^k)\| \\ &= \|(v^k - x) - \beta_k(z^k - x) - (1 - \beta_k)(y^k - x) \\ &\quad + (1 - \alpha_{k+1})((z^k - x) - (y^k - x))\| \leq \|v^k - x\| + \beta_k\|z^k - x\| \\ &\quad + (1 - \beta_k)\|y^k - x\| + (1 - \alpha_{k+1})\|z^k - x\| + (1 - \alpha_k)\|y^k - x\| \\ &\leq (C_S + C_x) + (C_T + C_x) + C_x + (C_T + C_x) + C_x = C_S + 2C_T + 5C_x. \end{aligned} \quad (52)$$

Moreover,

$$\begin{aligned} \|x^1 - x^0\| &= \|(x^1 - x) - (x^0 - x)\| \leq \|x^1 - x\| + \|x^0 - x\| \\ &\leq 2C_x \leq C_S + 2C_T + 5C_x, \end{aligned}$$

therefore all hypotheses of Lemma 24 are hold. Hence, the rate of convergence  $\{\|x^{k+1} - x^k\|\}$  is immediately implied by setting  $a_k = \|x^k - x^{k-1}\|$ ,  $b_k = \alpha_k$ ,  $\gamma = 1 - r$

and  $c_k$  as (52). By the following arguments, the rate for  $\{\|z^k - x^{k+1}\|\}$  can be derived

$$\begin{aligned}
 \|z^k - x^k\| &= \|z^k - (\alpha_k v^k + (1 - \alpha_k)\beta_k y^k + (1 - \alpha_k)(1 - \beta_k)z^k)\| \\
 &= \|\alpha_k(z^k - v^k) + (1 - \alpha_k)\beta_k(z^k - y^k) + (1 - \alpha_k)(1 - \beta_k)(z^k - z_k)\| \\
 &\leq \alpha_k\|v^k - z^k\| + (1 - \alpha_k)\beta_k\|z^k - y^k\| = \alpha_k\|v^k - z^k\| + \alpha_k\|z^k - y^k\| \\
 &= \alpha_k(\|v^k - z^k\| + \|z^k - y^k\|) \leq \frac{2}{(1-r)k} (C_S + C_T + 4C_x),
 \end{aligned} \tag{53}$$

where we have used the fact that  $(1 - \alpha_k)\beta_k \leq \alpha_k$ , and we also have

$$\begin{aligned}
 \|z^k - x^{k-1}\| &\leq \|z^k - x^k\| + \|x^k - x^{k-1}\| \\
 &\leq \frac{2}{(1-r)k} (C_S + C_T + 4C_x) + \frac{(C_S + 2C_T + 5C_x)J}{(1-r)k} \\
 &\leq \frac{(C_S + 2C_T + 5C_x)(J+2)}{(1-r)k}.
 \end{aligned}$$

□

### 3.2 The main result

Now, we are in a position to conclude our main result concerning the rate of convergence for convex trilevel optimization. Having proved that  $\|z^k - x^{k-1}\| \rightarrow 0$  as  $k \rightarrow \infty$  and considering the lower semi-continuity of  $\phi_1$ , one obtains that  $\{\phi_1(z^k)\}_{k \in \mathbb{N}}$  converges to the optimal value. Furthermore, this implies the convergence of the sequence  $\{\phi_1(x^k)\}_{k \in \mathbb{N}}$  to the same value.

We note that the same argument holds for the sequence  $\{\phi_2(z^k)\}_{k \in \mathbb{N}}$ . The following theorem presents the convergence rate in function values to their optima:

**Theorem 29** *Let  $\{x^k\}$ ,  $\{v^k\}$ ,  $\{z^k\}$  and  $\{y^k\}$  be sequences generated by Algorithm (1), where  $\alpha_k$  is proposed by (48). Then*

$$\begin{aligned}
 &\phi_1(z^{k+1}) - \phi_1(x^*) \\
 &\leq \frac{(J+2)(C_S + 2C_T + 5C_{x^*})\sqrt{JC_{x^*}(C_T + C_{x^*})}}{s\sqrt{1-r}(1-r)(k+1)\sqrt{k}} \quad \forall (s, k) \in \left(0, \frac{1}{L_{f_2}}\right] \times \mathbb{N},
 \end{aligned} \tag{54}$$

where  $C_S, C_T, C_{x^*}$  are the same constants as in Lemma 6,  $J$  is defined in (48), Furthermore, one has

$$\phi_2(y^{k+1}) - \phi_2(x^*) \leq \frac{C_{x^*} J}{2t(1-r)k}, \quad \forall (t, k) \in \left(0, \frac{1}{L_{f_1}}\right] \times \mathbb{N} \quad (55)$$

$$\phi_i(x^k) - \phi_i(x^*) \leq \sqrt{\frac{4(C_T + C_{x^*})J}{C_{x^*}(1-r)k}} \quad \text{for } i = 1, 2 \text{ and } k \in \mathbb{N} \quad (56)$$

$$\omega(x^k) - \omega(x^*) \leq \left(L_\omega + \frac{L_\omega^2}{2\mu} r C_{x^*}\right) r \sqrt{\frac{C_{x^*}(C_T + C_{x^*})J}{\gamma k}}. \quad (57)$$

**Proof** Since  $x^*$  is a solution of tri-level Problem 2 of Theorem 12, the following result was obtained:

$$\begin{aligned} \|x^{k+1} - x^*\|^2 &\leq (1 - (1-r)\alpha_k) \|x^k - x^*\|^2 + 2(1 - \alpha_k)\beta_k \langle T(x^*) \\ &\quad - x^*, x^{k+1} - x^* \rangle. \end{aligned} \quad (58)$$

Let us take

$$a_k = \|x^k - x^*\|^2, \quad c_k = \|\langle T(x^*) - x^*, x^{k+1} - x^* \rangle\|$$

and  $\alpha_k, \beta_k$  be as in (48). From (58), it follows that

$$\begin{aligned} a_{k+1} &\leq (1 - (1-r)\alpha_k) a_k + 2(1 - \alpha_k)\beta_k c_k \\ &\leq (1 - (1-r)\alpha_{k+1}) a_k + (\alpha_k - \alpha_{k+1}) c_k \end{aligned}$$

where we used the facts  $2(1 - \alpha_k)\beta_k \leq \alpha_k - \alpha_{k+1}$ , and  $\alpha_{k+1} \leq \alpha_k$ . By Lemma 6, we then have  $c_k \leq (C_T + C_{x^*})C_{x^*}$ . By utilizing Lemma 24, we obtain

$$\|x^k - x^*\|^2 \leq \frac{C_{x^*}(C_T + C_{x^*})J}{(1-r)k}, \quad \forall k \in \mathbb{N}. \quad (59)$$

Let us consider the first assertion (54). According to Proposition 26 and  $z^{k+1} = W(x^k)$ , for every step-size  $s \leq \frac{1}{L_{f_1}}$  the following inequality holds

$$\phi_1(z^{k+1}) - \phi_1(x^*) \leq \frac{1}{s} \langle x^k - z^{k+1}, x^k - x^* \rangle - \frac{1}{2s} \|x^k - z^{k+1}\|^2 \quad (60)$$

$$\leq \frac{1}{s} \langle x^k - z^{k+1}, x^k - x^* \rangle. \quad (61)$$

Combining with (51) and (59) for  $x^* \in X^* \subseteq \text{Fix}(W) = Y^*$ , one obtains

$$\begin{aligned} \langle x^k - z^{k+1}, x^k - x^* \rangle &\leq \|x^k - z^{k+1}\| \cdot \|x^k - x^*\| \\ &\leq \frac{(J+2)(C_S + 2C_T + 5C_{x^*})\sqrt{JC_{x^*}(C_T + C_{x^*})}}{\sqrt{1-r}(1-r)(k+1)\sqrt{k}}. \end{aligned} \quad (62)$$

Thus, the assertion (54) follows from (60) and (62).

Now, we obtain the rate of convergence for  $\phi_1(y^k) - \phi_1(x^*)$ . In Algorithm 1, we have  $y^{k+1} = T(x^k)$  and so using Lemma 25, one gets that

$$\phi_1(y^k) - \phi_1(x^*) = \phi_1(T(x^k)) - \phi_1(x^*) \leq \frac{1}{2t} \|x^k - x^*\|^2. \quad (63)$$

Plugging inequality (59) into (63) gives the desired assertion (55).

To establish the assertion (56), since  $\phi_i$ , ( $i = 1, 2$ ) is convex and bounded above on compact set  $\mathcal{B}(x^*, C_{x^*})$ , by invoking (Borwein and Vanderwerff 2010, Theorem 2.1.10) one concludes that  $\phi_i$  is Lipschitz on this set. Note thanks to  $x^k \in \mathcal{B}(x^*, C_{x^*})$ , one has

$$\phi_i(x^k) - \phi_1(x^*) \leq \frac{2}{C_{x^*}} \|x^k - x^*\| \quad (64)$$

which combined with (59) gives the desired assertion.

Finally, to bound the rate of convergence  $\omega(x^k) - \omega(x^*)$ , we take into consideration (49), (59), and strong convexity of  $\omega$ , for all  $k \in \mathbb{N} \setminus \{1\}$ . We obtain:

$$\begin{aligned} \omega(x^k) - \omega(x^*) &\leq \nabla\omega(x^*)^T (x^k - x^*) + \frac{1}{2\mu} \|\nabla\omega(x^k) - \nabla\omega(x^*)\|^2 \\ &\leq \left( L_\omega + \frac{L_\omega^2}{2\mu} \|x^k - x^*\| \right) \|x^k - x^*\| \\ &\leq \left( L_\omega + \frac{L_\omega^2}{2\mu} r C_{x^*} \right) r \sqrt{\frac{C_{x^*}(C_T + C_{x^*})J}{\gamma k}}. \end{aligned}$$

□

**Remark 30** It is worth pointing out that step-size  $\{\alpha_k\}$  depends on the parameter  $r$ , which needs to be chosen so that the map  $S$  is a contraction. Notice, however, that knowing  $L_\omega$  and  $\mu$ , one can consider  $r$  such that  $r \in (0, \frac{2}{L_\omega + \mu}]$ . In this case, the map  $S$  is guaranteed to be a contraction.

## 4 An extension to multilevel optimization problems

In this section, we extend our results to a multilevel convex optimization problem wherein we have an arbitrary number of nested minimization problems:

$$\left\{ \begin{array}{l} \arg \min_{x \in X_N^*} \omega(x), \\ X_N^* = \arg \min_{x \in X_{N-1}^*} [f_N(x) + g_N(x)], \\ \cdot \\ \cdot \\ \cdot \\ X_2^* = \arg \min_{x \in X_1^*} [f_2(x) + g_2(x)], \\ X_1^* = \arg \min_{x \in \mathbb{R}^n} [f_1(x) + g_1(x)], \end{array} \right.$$

It is then natural to define the following algorithm

$$x^{k+1} = \alpha_k^{(0)} S(x^k) + \alpha_k^{(1)} T_1(x^k) + \dots + \alpha_k^{(N)} T_N(x^k), \quad (65)$$

in which  $T_1, T_2, \dots, T_N$  are computed as the corresponding equation, i.e., (1), i.e.,

$$T_i(x) = \text{prox}_{t_i g_i}[x - t_i \nabla f_i(x)], \quad \forall i \in \{1, 2, \dots, N\},$$

and where the step-sizes  $(\alpha_k^{(0)}, \alpha_k^{(1)}, \alpha_k^{(2)}, \dots, \alpha_k^{(N)})$  satisfy the following:

- (P<sub>1</sub>) for all  $k \in \mathbb{N}$ , one has  $\sum_{j=0}^N \alpha_k^{(j)} = 1$ .
- (P<sub>2</sub>)  $\lim_{k \rightarrow \infty} \alpha_k^{(0)} = 0$ ,  $\sum_{k=1}^{+\infty} \alpha_k^{(0)} = \infty$  and,  $\limsup_k \frac{1}{\alpha_k^{(0)}} \sum_{i=2}^N \alpha_k^{(i)} = \delta^* \in [0, +\infty)$ .
- (P<sub>3</sub>)  $X_N^* \neq \emptyset$ ,  $\lim_{k \rightarrow \infty} \alpha_k^{(1)} = 1$  and for all  $j \in \{0, 2, \dots, N\}$  one has  $\lim_{k \rightarrow \infty} \alpha_k^{(j)} = 0$ .

The following facts hold:

- $X_N^* \subseteq X_{N-1}^* \subseteq \dots \subseteq X_1^* = \text{Fix}(T_1)$ .
- for every  $t_i \in (0, \frac{1}{L_{f_i}}]$  and  $i \in \{1, 2, \dots, N\}$  one has  $\tilde{X}_i^* = \text{Fix}(T_i)$  where

$$\tilde{X}_i^* = \arg \min_{x \in \mathbb{R}^n} [f_i(x) + g_i(x)].$$

- for every  $x \in VI(T_{i+1}, \text{Fix}(T_i))$  one has

$$\phi_i(T_{i+1}(x)) \leq \phi_i(y) \quad \forall y \in \text{Fix}(T_i),$$

where  $\phi_i(x) := f_i(x) + g_i(x)$  for all  $i \in \{1, 2, \dots, N\}$ .

- for each  $i \in \{1, 2, \dots, N-1\}$  one has  $X_{i+1}^* \cap \text{Fix}(T_{i+1}) = \text{Fix}(T_{i+1}) \cap \text{Fix}(T_i)$

- for  $i \in \{1, 2, \dots, N\}$ , one has

$$\phi_i(T_i(x)) - \phi_i(y) \leq \frac{1}{2t} \|x - y\|^2, \quad \forall x, y \in \mathbb{R}^n, \quad \forall t \in \left(0, \frac{1}{L_{f_i}}\right], \quad (66)$$

this follows from Lemma 25.

- for every  $x \in \text{dom}(\partial\phi_i)$  one has

$$\|x - T_i(x)\| \leq d(0, \partial\phi_i),$$

where  $\text{dom}(\partial\phi_i) = \{x \in \mathbb{R}^n : \partial\phi_i(x) \neq \emptyset\}$ .

**Assumption 31** (*Quadratic growth condition*) Suppose now that  $\phi_1$  grows quadratically (globally) away from a part of its minimizing set  $X_1^* = \text{Fix}(T_1)$ , i.e.,  $X_N^*$ , meaning there is a real number  $\eta > 0$  such that

$$\phi_1(x) \geq \phi_1^* + \frac{\eta}{2} \text{dist}^2(x, X_N^*) \quad \forall x \in \Omega^* \setminus \text{Fix}(T_1) \quad (67)$$

where  $\Omega^* = \mathcal{B}(x_0, C_{x_0})$  for given  $x_0 \in \text{Fix}(T_1)$  and  $\phi_1^*$  represents the optimal value of  $\phi_1$ .

In view of Fact 10 the following general result, however, holds true in having the qualification condition of being a non-empty interior  $X_N^*$ . We omit the proof.

**Lemma 32** If  $X_N^*$  has non-empty interior then  $X_N^* = \bigcap_{i=1}^N \text{Fix}(T_i)$ .

**Lemma 33** Let Assumption  $(P_2)$  hold. Then the sequence  $\{x^k\}$  generated by algorithm (65) is bounded.

**Proof** Define

$$R_k(x) = \frac{\alpha_k^{(1)}}{1 - \alpha_k^{(0)}} T_1(x) + \frac{\alpha_k^{(2)}}{1 - \alpha_k^{(0)}} T_2(x) + \dots + \frac{\alpha_k^{(N)}}{1 - \alpha_k^{(0)}} T_N(x), \quad \forall k \in \mathbb{N}.$$

Then, one can rewrite the sequence  $\{x^{k+1}\}$  as

$$x^{k+1} = \alpha_k^{(0)} S(x^k) + (1 - \alpha_k^{(0)}) R_k(x^k),$$

and since  $R_k$  is a convex combination of non-expansive operators, then it is non-expansive. Now, as in Lemma 6, for every  $x \in \text{Fix}(T_1)$  one has

$$\|x^{k+1} - x\| \leq \max \left\{ \|x^{k_0} - x\|, \frac{1}{1-r} (\|S(x) - x\| + \delta^* \|T_j(x) - x\|) \right\},$$

in which

$$\|T_j(x) - x\| = \max_{2 \leq i \leq N} \|T_i(x) - x\|$$

□



**Lemma 34** (Bauschke et al. 2011) *Let  $T_1, T_2, \dots, T_N$  be non-expansive mappings from  $\mathbb{R}^n$  to itself such that  $\cap_{i=1}^N T_i$  is non-empty and let  $\lambda_1, \lambda_2, \dots, \lambda_n$  be real numbers such that  $\sum_{i=1}^N \lambda_i = 1$ . Then*

$$\text{Fix} \left( \sum_{i=1}^N \lambda_i T_i \right) = \bigcap_{i=1}^N \text{Fix}(T_i).$$

**Lemma 35** *Let Assumption  $(P_2)$  hold. Then for every  $x \in \text{Fix}(T_1)$ , one has*

$$\begin{aligned} \|x^{k+1} - x\|^2 &\leq (1 - (1 - r)\alpha_k^{(0)})\|x^k - x\|^2 + \alpha_k^{(0)} \left\langle S(x) - x, x^{k+1} - x \right\rangle \\ &\quad + \sum_{i=2}^N \alpha_k^{(i)} \left\langle T_i(x) - x, x^{k+1} - x \right\rangle. \end{aligned}$$

**Proof** Suppose that  $x \in \text{Fix}(T_1)$  is given. First, consider

$$\begin{aligned} C_k &:= \alpha_k^{(0)}(S(x^k) - S(x)) + \alpha_k^{(1)}(T_1(x^k) - T_1(x)) + \dots \\ &\quad + \alpha_k^{(N)}(T_N(x^k) - T_N(x)), \\ \text{and} \\ D_k &:= \alpha_k^{(0)}(S(x) - x) + \alpha_k^{(2)}(T_2(x) - x) + \dots + \alpha_k^{(N)}(T_N(x) - x). \end{aligned}$$

Therefore, we have

$$C_k + D_k = x^{k+1} - x, \quad \|C_k\| \leq (1 - (1 - r)\alpha_k^{(0)})\|x^k - x\|. \quad (68)$$

To see the second part, for all  $k \in \mathbb{N}$  we have

$$\begin{aligned} \|C_k\| &\leq r\alpha_k^{(0)}\|x^k - x\| + \alpha_k^{(1)}\|x^k - x\| + \dots + \alpha_k^{(N)}\|x^k - x\| \\ &= (r\alpha_k^{(0)} + \alpha_k^{(1)} + \dots + \alpha_k^{(N)})\|x^k - x\| \\ &= (1 - (1 - r)\alpha_k^{(0)})\|x^k - x\|. \end{aligned}$$

Now, by plugging (68) in the following inequality

$$\|C_k + D_k\|^2 \leq \|C_k\|^2 + \langle D_k, C_k + D_k \rangle,$$

the assertion follows immediately.  $\square$

Now, we are in a position to present our main result regarding multi-level scenarios.

**Theorem 36** *Assume that  $X^*$  has non-empty interior. Moreover the following holds*

$$(A_1^*) \limsup_k \frac{\sum_{i=1}^N \|\alpha_k^{(i)} - \frac{1 - \alpha_k^{(0)}}{N}\|}{\alpha_k^{(0)}} = 0.$$

Then  $\{x^k\}$  convergence to some unique  $x_N^*$  such that

$$\langle x_N^* - S(x_N^*), y - x_N^* \rangle \geq 0, \quad \forall y \in X_N^*.$$

**Proof** First, we note that by looking at Lemma 32 we observe that

$$X^* = \text{Fix} \left( \frac{\sum_{i=1}^N T_i}{N} \right).$$

Let us now, consider the auxiliary sequence

$$y^{k+1} = \alpha_k^{(0)} S(y^k) + (1 - \alpha_k^{(0)}) \frac{\sum_{i=1}^N T_i(y^k)}{N}.$$

By employing (Xu 2004, Theorem 3.2) along with Lemma 34, one establishes the convergence of the sequence  $y^k$  to a specific point denoted as  $x_N^*$ . Consequently, we deduce the following:

$$\begin{aligned} \|x^{k+1} - y^{k+1}\| &\leq r\alpha_k^{(0)} \|x^k - y^k\| + \left\| \sum_{i=1}^N \alpha_k^{(i)} T_i(x^k) - (1 - \alpha_k^{(0)}) \frac{\sum_{i=1}^N T_i(y^k)}{N} \right\| \\ &\leq (1 - (1 - r)\alpha_k^{(0)}) \|x^k - y^k\| + \left\| \sum_{i=1}^N (\alpha_k^{(i)} - \frac{1 - \alpha_k^{(0)}}{N}) T_i(y^k) \right\| \\ &\leq (1 - (1 - r)\alpha_k^{(0)}) \|x^k - y^k\| + \left( \sum_{i=1}^N \left\| \alpha_k^{(i)} - \frac{1 - \alpha_k^{(0)}}{N} \right\| \right) \|T_i(y^k)\| \end{aligned}$$

using Lemma 5 follows that  $\|x^k - y^k\|$  goes to zero as  $k \rightarrow \infty$ .  $\square$

**Theorem 37** Let Assumption  $(P_2)$  hold with  $\delta^* = 0$ , together with 31. Then the sequence  $\{x^k\}$  generated by algorithm (65) converges to some unique  $x_N^* \in X_N^*$  such that

$$\langle x_N^* - S(x_N^*), y - x_N^* \rangle \geq 0, \quad \forall y \in X_N^*,$$

and

$$\min_{x \in X_N^*} \omega(x) = \omega(x_N^*).$$

**Proof** Let  $x_N^* \in X_N^*$  be the unique fixed point of the contraction  $P_{X_N^*} S$ , namely the unique solution of  $VI(S, X_N^*)$ , i.e.,

$$\langle x_N^* - S(x_N^*), x - x_N^* \rangle \geq 0, \quad \forall x \in X_N^*. \quad (69)$$

Invoking Assumption  $(P_2)$ , one gives that  $\{x^k\}$  is bounded, and so  $x^k \in \Omega^*$ . Furthermore, utilizing assumption 31, it is straightforward to show that  $w(x^k) \subset X_N^*$ . Moreover, one can extract a convergent sub-sequence  $\{x^{k_i}\}$  of  $\{x^{k+1}\}$  or any sub-sequence thereof to  $x_N' \in X_N^*$ , which holds by Lemma 8, part c, and (69) so that

$$\begin{aligned} \limsup_k \langle S(x_N^*) - x_N^*, x^{k+1} - x_N^* \rangle &= \lim_i \langle S(x_N^*) - x_N^*, x^{k_i} - x_N^* \rangle \\ &= \langle S(x_N^*) - x_N^*, x_N' - x_N^* \rangle \leq 0. \end{aligned} \quad (70)$$

and now, using Lemma 35 follows that

$$\begin{aligned} \|x^{k+1} - x_N^*\|^2 &\leq (1 - (1 - r)\alpha_k^{(0)})\|x^k - x_N^*\|^2 + \alpha_k^{(0)} \langle S(x_N^*) - x_N^*, x^{k+1} - x_N^* \rangle \\ &\quad + \sum_{i=2}^N \alpha_k^{(i)} \langle T_i(x_N^*) - x_N^*, x^{k+1} - x_N^* \rangle. \end{aligned} \quad (71)$$

Next, set

$$\begin{cases} a_k = \|x^k - x^*\|^2, \\ \gamma_k = (1 - r)\alpha_k^{(0)}, \\ \delta_k = \alpha_k^{(0)} \langle S(x_N^*) - x_N^*, x^{k+1} - x_N^* \rangle + \sum_{i=2}^N \alpha_k^{(i)} \langle T_i(x_N^*) - x_N^*, x^{k+1} - x_N^* \rangle. \end{cases} \quad (72)$$

One has that

$$a_{k+1} \leq (1 - \gamma_k)a_k + \delta_k.$$

Additionally, utilizing the boundedness of  $\{x^k\}$  together with  $\delta^* = 0$  it can be inferred that  $\limsup_k \frac{\delta_k}{\gamma_k} \leq 0$ . Indeed, taking into account (70) and  $\delta^* = 0$  gives

$$\begin{aligned} \limsup_k \frac{\delta_k}{\gamma_k} &= \limsup_k \left[ \frac{\alpha_k^{(0)} \langle S(x_N^*) - x_N^*, x^{k+1} - x_N^* \rangle + \sum_{i=2}^N \alpha_k^{(i)} \langle T_i(x_N^*) - x_N^*, x^{k+1} - x_N^* \rangle}{\alpha_k^{(0)}} \right] \\ &\leq \limsup_k \langle S(x_N^*) - x_N^*, x^{k+1} - x_N^* \rangle + C \limsup_k \frac{\sum_{i=2}^N \alpha_k^{(i)}}{\alpha_k^{(0)}} \leq 0 \end{aligned}$$

where  $C = \max_{1 \leq i \leq N} \sup_k \|T_i(x_N^*) - x_N^*\| \|x^{k+1} - x_N^*\|$ . The desired claim can now be deduced from Lemma 5.  $\square$

We will now show the rate of convergence for the general case. To study this, let us take the sequences  $y_k^{(i)} = T_i(x^{k-1})$  and step-sizes

$$\alpha_k^{(i)} = \min\left\{\frac{2}{(1-r)k}, 1\right\}, \quad \beta_k^{(i)} = \frac{\alpha_k^{(i)} - \alpha_{k+1}^{(i)}}{2(2 - \alpha_k^{(i)})}, \quad J = \lfloor \frac{2}{1-r} \rfloor \quad (73)$$

**Theorem 38** Let  $y_k^{(i)} = T_i(x^{k-1})$  be sequences generated by Algorithm (65), where  $\alpha_k$  is proposed by (73). Then

$$\begin{aligned} & \phi_1(y_k^{(1)}) - \phi_1(x_N^*) \\ & \leq \frac{(J+2)(C_S + 2C_{T_1} + 5C_{x_N^*})\sqrt{JC_{x_N^*}(C_{T_1} + C_{x_N^*})}}{s\sqrt{1-r}(1-r)(k+1)\sqrt{k}} \quad \forall (s, k) \in \left(0, \frac{1}{L_{f_2}}\right] \times \mathbb{N}, \end{aligned} \quad (74)$$

where  $C_S, C_{T_1}, C_{x_N^*}$  are the same constants as in Lemma 6 from Lemma 33,  $J$  is defined in, Furthermore, one has

$$\phi_i(y_k^{(i)}) - \phi_i(x_N^*) \leq \frac{C_{x_N^*}J}{2t(1-r)k}, \quad \forall (t, k) \in \left(0, \frac{1}{L_{f_i}}\right] \times \mathbb{N}, \quad i \neq 1 \quad (75)$$

$$\omega(x^k) - \omega(x_N^*) \leq \left(L_\omega + \frac{L_\omega^2}{2\mu}rC_{x^*}\right)r\sqrt{\frac{C_{x_N^*}(C_{T_1} + C_{x_N^*})J}{\gamma k}}. \quad (76)$$

**Proof** As (59) and Lemma 35 and utilizing of Lemma 24 one can conclude that

$$\|x^k - x_N^*\| \leq \frac{C_{x_N^*}(C_{T_1} + C_{x_N^*})J}{(1-r)k},$$

and the rest of the proof is similar to the one for the trilevel Theorem 29.  $\square$

## 5 Conclusion

We have shown how to approach a broad class of hierarchical convex optimization problems wherein the inner problems optimize the so-called composite functions, i.e., sums of a convex smooth function and a convex non-smooth one, and all but the inner-most problem consider a constraint set composed of minimizers of another problem. We have used proximal gradient operators in an iterative proximal-gradient algorithm related to “SAM” of Sabach and Shtern (2017). For the first time, we consider diminishing sequences  $\alpha_k$  and  $\beta_k$  such that the large limit of  $\frac{\beta_k}{\alpha_k}$  need not exist. The convergence is studied in a number of cases, depending on the relative speed of convergence of  $\alpha_k$  and  $\beta_k$  and in some cases regularity properties of the problem layers.

We showed standard  $\mathcal{O}(\sqrt{\frac{1}{k}})$ ,  $\mathcal{O}(\frac{1}{k})$ ,  $\mathcal{O}(\frac{1}{(k+1)\sqrt{k}})$  rates of convergence for appropri-

ate corresponding quantities. Future work can include introducing stochasticity to the problems.

**Acknowledgements** Shortly after we have posted our first draft on-line in arxiv, a team from the The University of Tokyo, RIKEN, and The Institute of Statistical Mathematics have submitted their draft (Sato et al. 2021), which considered rather a different method for a closely related problem, albeit without bounding rates of convergence. This work has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101070568.

**Funding** Open access publishing supported by the National Technical Library in Prague. The research leading to these results received funding from OP RDE under Grant Agreement No CZ.02.1.01/0.0/0.0/16\_019/0000765. This work was supported by Czech Science Foundation (Grant number 22-15524 S).

## Declarations

**Conflict of interest** The authors have no conflicts of interest to declare that are relevant to the content of this article.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Al-Khayyal F, Horst R, Pardalos PM (1992) Global optimization of concave functions subject to quadratic constraints: an application in nonlinear bilevel programming. *Ann Oper Res* 34:125–147
- Beck A, Teboulle M (2009) A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J Imaging Sci* 2:183–202
- Beck A, Sabach S (2014) A first order method for finding minimal norm-like solutions of convex optimization problems. *Math Program* 147:25–46
- Bialas WF, Karwan MH (1973) Mathematical programs with optimization problems in the constraints. *Oper Res* 21:37–44
- Bauschke H, Combettes H, Patrick L (2011) *Convex analysis and monotone operator theory in Hilbert spaces*. Springer, New York
- Ben-Tal Aharon, El Ghaoui Laurent, Nemirovski Arkadi (2009) *Robust optimization*, vol 28. Princeton University Press, Princeton
- Borwein JM, Guoyin L, Matthew T (2017) Convergence rate analysis for averaged fixed point iterations in common fixed point problems. *SIAM J Optim* 27:1–33
- Blair C (1992) The computational complexity of multi-level linear programs. *Ann Oper Res* (34)
- Borwein JM, Vanderwerff J (2010) *Convex functions: constructions, characterizations, and counterexamples*. Cambridge University Press, Cambridge (172)
- Bolte J, Daniilidis A, Lewis A (2007) The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. *SIAM J Optim* 17(4):1205–1223
- Bauschke H, Noll D, Phan HM (2015) Linear and strong convergence of algorithms involving averaged nonexpansive operators. *J Math Anal Appl* 421(1):1–20
- Dempe S, Dutta J, Mordukhovich BS (2007) New necessary optimality conditions in optimistic bilevel programming. *Optimization* 56:577–604
- Dempe S, Dutta J, Mordukhovich BS (2014) Necessary optimality conditions in pessimistic bilevel programming. *Optimization* 56:505–533

- Dempe S, Dutta J, Mordukhovich BS (1994) An iterative method for solving a class of nonlinear operator equations in Banach space. *J Panamerican Math* 4:39–54
- Fisac Jaime F, Sastry S, Shankar (2015) The pursuit-evasion-defense differential game in dynamic constrained environments. In: 2015 54th IEEE conference on decision and control (CDC), 4549–4556
- Iiduka H (2011) Iterative algorithm for solving triple-hierarchical constrained optimization problem. *J Optim Theory Appl* 148:580–592
- Isao Y (2001) The hybrid steepest descent method for variational inequality problems over the intersection of the fixed-point sets of nonexpansive mappings, inherently parallel algorithms in feasibility and optimization and their applications, Edited by D. Butnariu, Y. Censor, and S. Reich, North-Holland, Amsterdam, Holland pp 473–504
- Lampariello L, Neumann C, Ricci J, Sagratella S, Stein O (2020) An explicit Tikhonov algorithm for nested variational inequalities. *J Comput Appl* 77:335–350
- Lampariello L, Sagratella S, Stein O (2019) The standard pessimistic bilevel problem. *SIAM J Optim* 29:1634–1656
- Lu XW, Xu HK, Yin XM (2009) Hybrid methods for a class of monotone variational inequalities. *Nonlinear Anal* 71:1032–1041
- Moudafi A (2007) Krasnoselski–Mann iteration for hierarchical fixed-point problems. *Inverse Probl* 23:1635–1640
- Maingé PE, Abdellatif M (2007) Strong convergence of an iterative method for hierarchical fixed-point problems. *Pac J Optim* 3:529–538
- Nesterov Y (2003) *Introductory lectures on convex optimization* (vol 87). Springer, Applied Optimization (APOP)
- Solodov M (2007) An explicit descent method for bilevel convex optimization. *J Convex Anal* 14:277
- Sabach S, Shtern S (2017) A first order method for solving convex bilevel optimization problems. *SIAM J Optim* 27:640–660
- Sato R, Mirai T, Takeda A (2021) A gradient method for multilevel optimization. *Adv Neural Inf Process Syst* 34:7522–7533
- Senter HF, Dotson (1974) Approximating fixed points of nonexpansive mappings. *Proc Am Math Soc* 44:375–380
- Wilfred W (1977) *Multi-level programming*. World Bank
- Xu HK (2010) Viscosity method for hierarchical fixed point approach to variational inequalities. *Taiwan J Math* 14:463–478
- Xu HK (2004) Viscosity approximation methods for nonexpansive mappings. *J Math Anal Appl* 298:279–291
- Xu HK (2002) Iterative algorithms for nonlinear operators. *J Lond Math Soc* 66:240–256
- Zhang R (1994) Problems of hierarchical optimization in finite dimensions. *SIAM J Optim* 4:521–536

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.