

NESTEROV PERTURBATIONS AND PROJECTION METHODS APPLIED TO IMRT

ESTHER BONACKER¹, AVIV GIBALI^{2,3,*}, KARL-HEINZ KÜFER¹

¹*Optimization Department, The Fraunhofer ITWM, 67663 Kaiserslautern, Germany*

²*Department of Mathematics, ORT Braude College, Karmiel 2161002, Israel*

³*The Center for Mathematics and Scientific Computation, University of Haifa, Mt. Carmel, Haifa 3498838, Israel*

Abstract. In this paper, we use the simultaneous and cyclic subgradient projection methods with perturbations and apply them to linear and nonlinear optimization problems arising naturally in IMRT. The numerical results illustrate the mechanism of the perturbations and demonstrate that they lead to a significantly faster termination of the algorithm than the unperturbed projection methods, especially when the problem is ill-conditioned.

Keywords. Projection methods; Subgradient; Bounded perturbation resilient; Inertial methods.

1. INTRODUCTION

Consider the following *general constrained convex optimization problem*

$$\begin{aligned} \text{Minimize}_{x \in \mathbb{R}^n} & f(x) \\ \text{s.t. } & g_j(x) \leq 0, \quad j = 1, \dots, m, \end{aligned} \tag{1.1}$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R}$ and $g_j : \mathbb{R}^n \rightarrow \mathbb{R}$ for $j \in J = \{1, \dots, m\}$ are convex functions. We further assume that f is bounded from below to ensure the existence (if not uniqueness) of a solution x^* of (1.1).

In order to solve this problem, Gibali *et al.* [23] proposed to translate the model into an equivalent epigraph form (2.1) and then apply the so-called *level set scheme*, which constructs a sequence of convex feasibility problems. The general principle of this procedure is to sequentially bound the function f from above and then use its level set as a new additional convex constraint to the feasibility problem. A suitable algorithmic operator T is then used to solve the feasibility problem, and the upper bound on f is lowered and the process continues.

As shown in [23], the result of the level set scheme x_ε^* approximates x^* (a solution of (1.1)) in the sense that $f(x_\varepsilon^*) < f(x^*) + \varepsilon$ for $\varepsilon > 0$, and $g_j(x_\varepsilon^*) \leq 0$ for all $j \in J$. x_ε^* is called an ε -optimal solution of (1.1). The value of ε depends on the update strategy for the upper bound on f , which we discuss later.

*Corresponding author.

E-mail addresses: bonacker@itwm.fraunhofer.de (E. Bonacker), aviv@braude.ac.il (A. Gibali), kuefer@itwm.fraunhofer.de (K.-H. Küfer).

Received November 26, 2019; Accepted February 4, 2020.

For solving (1.1) via the level set scheme, we choose the *simultaneous* and *cyclic subgradient projection* methods as operator T to solve the arising feasibility problems. Both methods are bounded perturbation resilient and [2, Theorem 11 and Corollary 14] guarantees that the convergence rate of the unperturbed methods is preserved. We introduce new perturbations in order to solve the feasibility problems within even fewer iterations. We will formulate them both as inner and outer perturbations, see, e.g., [16, 20, 34].

The most well-known approach to incorporate perturbations is the *superiorization methodology* which uses perturbations pro-actively during the performance of the iteration scheme to steer the algorithm to an output that still solves the given problem, but is superior with respect to a given secondary criterion; see, e.g., [5–7, 9, 10, 13]. This approach has been used in the same application as the one we consider in this paper: Intensity-modulated radiation therapy (IMRT) treatment planning [4, 23].

However, our motivation to use perturbations is a different one. In IMRT problems, a zigzagging behavior of the projection methods is observed, which slows their progress down. This happens due to the fact that we attempt to satisfy two main goals, which conflict with each other: Irradiating the tumor and at the same time sparing surrounding healthy organs. Our goal is to use perturbations to address this behavior and accelerate the progress of the projection methods. In [3], we did this by constructing new non-zigzagging directions and using those directions as perturbations; see also [37]. Clearly there exist many optimization techniques such as [18, 21, 22, 31], but here we mainly focus on first order methods with perturbations.

Following the recent results in [3], in this work, we introduce new inner and outer perturbations, called *Nesterov perturbations*, in order to address the zigzagging behavior in a new way. These perturbations are then plugged into the level set scheme for solving (1.1) with the simultaneous and cyclic subgradient projection methods as operator T . In contrast to the perturbations used in [3], the Nesterov perturbations do not generate a new non-zigzagging direction. Instead they enlarge the steps into the directions generated by T itself and thereby amplify the movement towards the feasible set, which is already contained in those directions.

The outline of the paper is as follows. We start in Section 2 by presenting the mathematical background needed for introducing our results. In Section 3, we introduce the perturbation developed by us and show that it is bounded. In Section 4, we present a linear feasibility problem on which we demonstrate the main numerical effects of the perturbed projection methods. Finally we apply the perturbed projection methods to nonlinear optimization problems arising from IMRT treatment planning and discuss the influence of several parameters on their performance.

2. MATHEMATICAL BACKGROUND

Consider the general constrained convex optimization problem (1.1). Its equivalent *epigraph form* is defined as follows

$$\begin{aligned} \text{Minimize}_{x \in \mathbb{R}^n, t \in \mathbb{R}} \quad & t \in \mathbb{R} \\ \text{s.t.} \quad & f(x) - t \leq 0 \\ & g_j(x) \leq 0, \quad j = 1, \dots, m \end{aligned} \tag{2.1}$$

with the optimal value denoted by $t^* \in \mathbb{R}$ and there exists a point x^* with $f(x^*) = t^*$, which is feasible for (2.1). Now, (2.1) is transformed into a sequence of *convex feasibility problems*

(CFP) of the form:

$$\text{Find } x \in C := \bigcap_{i \in I} C_i, \quad (2.2)$$

where $C_i := \{x \in \mathbb{R}^n \mid \varphi_i(x) \leq 0\}$ and the functions φ_i are convex for all $i \in I$.

Now we describe the so-called *level set scheme* [23] which uses a decreasing sequence $\{t_s\}_{s=1}^{\infty}$. For any $s \geq 1$, define the functions

$$\varphi_1^s := f(x) - t_s, \varphi_{j+1}^s := g_j(x) \quad j = 1, \dots, m \quad (2.3)$$

and then define the convex feasibility problem P^s :

$$\text{Find } x \in C^s = \bigcap_{i \in I} C_i^s, \quad (2.4)$$

where $C_i^s := \{x \in \mathbb{R}^n \mid \varphi_i^s(x) \leq 0\}$ for all $i \in I = \{1, \dots, m+1\}$. If $C \neq \emptyset$, a solution x_s^* of P^s is found by a suitable algorithmic operator T (T can produce a solution in a finite number of iterations or approximate it up to a given accuracy based on the application). Next, the upper bound t_s is updated and t_{s+1} is determined according to some user-defined rule. Examples for this rule are (a) $t_{s+1} = f(x_s^*) - \varepsilon_s$ and (b) $t_{s+1} = f(x_s^*)(1 - \varepsilon_s)$, where $\varepsilon_s > 0$ for all $s \geq 0$. Using t_{s+1} the next convex feasibility problem, denoted by P^{s+1} is formulated and solved sequentially. If $C^s = \emptyset$, then x_{s-1}^* is the result of the level set scheme and a ε -optimal solution of (1.1), where $\varepsilon = \varepsilon_{s-1}$ for update strategy (a) and $\varepsilon = f(x_{s-1}^*)\varepsilon_{s-1}$ for strategy (b).

In general, it is non-trivial to determine if $C^s = \emptyset$. Therefore, we consider x_{s-1}^* to be the result of the level set scheme when the algorithmic operator T is unable to find a solution to P^s within a given number of iterations. There are many iterative methods for solving feasibility problems, and one particular class of methods is the class of projection methods; see, e.g., [11, 14, 15, 38].

We introduce two specific projection methods, which we use as the algorithmic operator T for the level set scheme. We start with the cyclic subgradient projection method.

Let $x^0 \in \mathbb{R}^n$ be an arbitrary starting point. Given the current iterate x^k , the next iterate x^{k+1} can be calculated via

$$x^{k+1} = T(x^k) = x^k + \lambda_k p(x^k) \quad (2.5)$$

with

$$p(x^k) = - \frac{\max\{0, \varphi_{i(k)}(x^k)\}}{\|\xi^k\|^2} \xi^k, \quad (2.6)$$

where $\xi^k \in \partial \varphi_{i(k)}(x^k)$ (subgradient of $\varphi_{i(k)}$ at x^k) is arbitrary, $\lambda_k \in [\varepsilon_1, 2 - \varepsilon_2]$ (relaxation parameters) for arbitrary $\varepsilon_1, \varepsilon_2 \in (0, 1]$ and $\{i(k)\}$ is a cyclic control sequence, that is, $i(k) = (k \bmod m) + 1$. Such methods belong to the class, which is also known as incremental subgradient projection algorithm or also ε -subgradient methods. For further information and results, the readers are referred to [1, 17, 19, 29, 32], just to name a few.

Our second option for choosing T is the *simultaneous subgradient projection* method.

Let $x^0 \in \mathbb{R}^n$ be an arbitrary starting point. Given the current iterate x^k , the next iterate x^{k+1} can be calculated via

$$x^{k+1} = T(x^k) = x^k + \lambda_k p(x^k) \quad (2.7)$$

with

$$p(x^k) := - \sum_{i \in I} w_i \frac{\max\{0, \varphi_i(x^k)\}}{\|\xi^k\|^2} \xi^k, \quad (2.8)$$

where $\xi^k \in \partial \varphi_i(x^k)$ is an arbitrary subgradient, $w_i > 0$ are weights with $\sum_{i \in I} w_i = 1$ and $\lambda_k \in [\varepsilon_1, 2 - \varepsilon_2]$ (relaxation parameters) for arbitrary $\varepsilon_1, \varepsilon_2 \in (0, 1]$.

In the following, we use the notation $\bar{p}(x^k) := p(x^k)/\|p(x^k)\|$, and refer to the simultaneous subgradient projection method simply as simultaneous projection and to the cyclic subgradient projection method as cyclic projection.

The convergence proof of the level set scheme [23, Theorem 3.6] relies on using the method that is either finite convergence, if there exists such a method for the given problem, or a solution in finite number of iteration is considered as an approximated solution with respect to some accuracy ε given in advance. We can transform both the cyclic and the simultaneous projections to fulfill this property by using the approach of [30]. Note that [23, Theorem 3.6], in contrast to convergence proofs for other finite convergent projection methods, e.g., [8], is not based on the assumption that the Slater Condition holds. A solution in finite number of iterations is considered an approximate solution with respect to some accuracy ε . In our context, this means that we do not rely on the existence of $x \in \mathbb{R}^n$ with $\varphi_i(x) < 0$ for all $i \in I$ for the level set scheme to converge.

Definition 2.1. Given a problem P , an algorithmic operator T and a starting point x^0 such that the sequence $\{x^k\}_{k=0}^\infty$ generated by $x^{k+1} = T(x^k)$ converges to a solution of P . Then T is called *bounded perturbation resilient* if any sequence $\{y^k\}_{k=0}^\infty$ with $y^0 = x^0$ generated using either *inner perturbations* via

$$y^{k+1} = T(y^k + \beta_k v^k) \quad (2.9)$$

for all iterations $k \geq 0$ or using *outer perturbations* via

$$y^{k+1} = T(y^k) + \beta_k v^k \quad (2.10)$$

for all iterations $k \geq 0$, where $\beta_k v^k$ are bounded perturbations (i.e., $\beta_k \in \mathbb{R}_{\geq 0}$ for all $k \geq 0$, $\sum_{k=0}^\infty \beta_k < \infty$, $v^k \in \mathbb{R}^n$ and $\|v^k\| \leq M \in \mathbb{R}$ for all $k \geq 0$) also converges to a solution of P .

Both the cyclic and the simultaneous projection method are known to be bounded perturbation resilient; see e.g. [12].

2.1. The inner and outer perturbation scheme. In this section, we describe the specific choices of operator T and sequences $\{\beta_k\}$ and $\{v^k\}$ that we refer to in the rest of our work.

Given a function $\lambda : \mathbb{R}^n \rightarrow [\varepsilon_1, 2 - \varepsilon_2]$ with arbitrary $\varepsilon_1, \varepsilon_2 \in (0, 1]$ (this function is used in place of the relaxation parameters λ_k in (2.5) or (2.7)). Let the algorithmic operator

$$T(x) = x + \lambda(x)p(x), \quad (2.11)$$

where $p(x)$ is either the simultaneous (2.8) or the cyclic projection step (2.6).

Let $c : \mathbb{R}^n \rightarrow \{\text{false}, \text{true}\}$ and $\tilde{c} : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \{\text{false}, \text{true}\}$ with

$$\tilde{c}(y^k, y^{k-1}) := \begin{cases} \neg \tilde{c}(y^{k-1}, y^{k-2}) \wedge c(y^k), & k > 1, \\ c(y^k), & k \leq 1 \end{cases} \quad (2.12)$$

be functions, which determine whether perturbation is applied in a certain iteration. $\tilde{c}(y^k, y^{k-1})$ is by construction only true, if $\tilde{c}(y^{k-1}, y^{k-2})$ is false and $c(y^k)$ is true, i.e., if iteration $(k-1) \rightarrow k$ has not been perturbed and the condition c is fulfilled at y^k .

The function c we use in our work is the following

$$c(y^k) = \begin{cases} \text{true}, & \text{if } \langle \bar{p}(y^{k-1}), \bar{p}(y^k) \rangle \in [-1 + \varepsilon_{\min}, -1 + \varepsilon_{\max}] \wedge k > 0 \\ \text{false}, & \text{otherwise,} \end{cases} \quad (2.13)$$

where $\bar{p}(y^k) = p(y^k)/\|p(y^k)\|$ and $0 \leq \varepsilon_{\min} < \varepsilon_{\max} \leq 2$. Choosing c like this means that an iteration is perturbed when the current projection step $\bar{p}(y^k)$ and the previous one $\bar{p}(y^{k-1})$ are strongly opposing. What exactly strongly opposing means is determined by the threshold parameters $\varepsilon_{\min}, \varepsilon_{\max}$, which result in a maximum (corresponding to ε_{\min}) and minimum (corresponding to ε_{\max}) angle between $\bar{p}(y^k)$ and $\bar{p}(y^{k-1})$.

Let

$$\beta_k = \begin{cases} 1, & k \leq K, \\ 0, & k > K \end{cases} \quad (2.14)$$

for some $K \in \mathbb{N}$. This specific choice of $\{\beta_k\}_{k=0}^{\infty}$ fulfills the general requirements given in Definition 2.1. It implies that in iterations $k > K$ the unperturbed operator T is used. In terms of convergence, this means that the perturbed and the unperturbed iteration scheme behave the same way. For a more general choice of $\{\beta_k\}_{k=0}^{\infty}$ (for which the properties required by Definition 2.1 hold) [2, Theorem 11 and Corollary 14] ensures that the convergence rate of the unperturbed method is preserved by its perturbed counterpart.

With regard to the numerical application of any method modified by bounded perturbations, (2.14) is a natural way of choosing $\{\beta_k\}_{k=0}^{\infty}$ because even if in theory all β_k were chosen as > 0 , they would be treated as 0 by the computer after some iteration index due to the limited machine precision. The second point we would like to raise about (2.14) is that, in practice, the algorithm to solve a given CFP is terminated either when a solution has been found or when a maximum number n_{\max} of iteration has been reached. Our goal is to use perturbations to find solutions to CFPs within fewer iterations. Thereby we are also able to find solutions with lower objective function values when the perturbed iteration scheme manages to find them within n_{\max} iterations while the unperturbed scheme does not. Like this, we achieve a faster numerical termination of the of the overall procedure. In line with this reasoning, we choose K as n_{\max} in our computations.

Let $\{y^k\}_{k=0}^{\infty}$ be the sequence of iterates produced by the *outer perturbation scheme*, which is defined as follows.

$$y^{k+1} = T(y^k) + \beta_k v^k, \quad (2.15)$$

where

$$v^k = \begin{cases} b^k - \lambda(y^k)p(y^k), & \text{if } \tilde{c}(y^k, y^{k-1}) = \text{true}, \\ 0, & \text{otherwise,} \end{cases} \quad (2.16)$$

and $\{b^k\}_{k=0}^{\infty}$ is bounded.

Let $\{z^k\}_{k=0}^{\infty}$ be the sequence of iterates produced by the *inner perturbation scheme*, which is defined as follows.

$$z^{k+1} = T(z^k + \beta_k v^k), \quad (2.17)$$

where

$$v^k = \begin{cases} b^k, & \text{if } c(z^k) = \text{true}, \\ 0, & \text{otherwise,} \end{cases} \quad (2.18)$$

and $\{b^k\}_{k=0}^{\infty}$ is bounded.

Proposition 2 in [3] implies that the inner and the outer perturbation scheme with β_k chosen as in (2.14) can be used interchangeably with regard to weak and strong convergence of the resulting sequences of iterates $\{y^k\}_{k=0}^{\infty}$ and $\{z^k\}_{k=0}^{\infty}$ when $y^0 = z^0$.

3. PERTURBATION OF THE SIMULTANEOUS AND CYCLIC SUBGRADIENT PROJECTION METHOD

Consider the unconstrained minimization problem:

$$\text{Minimize } f(x)$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is a convex function.

The Nesterov acceleration method [27] for solving the problem is phrased as follows. Choose starting points x_0 and $y_0 = x_0 \in \mathbb{R}^n$. Define

$$x_k = y_{k-1} - s \nabla f(y_{k-1}), \quad (3.1)$$

$$y_k = x_k + \frac{k-1}{k+2}(x_k - x_{k-1}). \quad (3.2)$$

When s is a fixed step size with $s < 1/L$, where L is the Lipschitz constant of ∇f , this method has a convergence rate of

$$f(x_k) - f(x^*) = \mathcal{O}\left(\frac{\|x_0 - x^*\|^2}{sk^2}\right),$$

where x^* is any minimizer of f . This convergence rate has been shown to be the optimal rate among all methods, which use only use gradient information of f at each iteration [26].

Since the introduction of Nesterov's scheme, there has been much work on the development of first-order accelerated methods, Theoretical developments can be found in [24–26] and [36] offers a unified analysis of these ideas.

Inspired by the above, we introduce new perturbations which we call the Nesterov perturbations. Observe that the update rule (3.1) for x_k is structurally similar to simultaneous projection in the sense that it uses only gradient information of y_{k-1} . We adopt the momentum term $y_k - x_k$ from (3.2) as perturbation vector b^k .

The inner perturbation scheme (2.17) using the Nesterov perturbation is

$$z^{k+1} = T(z^k + \beta_k b^k), \quad (3.3)$$

where

$$b^k = \begin{cases} \lambda_k^{NE} \frac{k-1}{k+2} (z^k - z^{k-1}), & \text{if } c(z^k) = \text{true}, \\ 0, & \text{otherwise,} \end{cases} \quad (3.4)$$

$\{\beta_k\}_{k=0}^\infty$ is chosen as in (2.14), $\{\lambda_k^{NE}\}_{k=0}^\infty$ is a user-chosen bounded sequence of step sizes with $\lambda_k^{NE} \in [0, \Lambda]$ for all $k \geq 0$ and c is defined as in (2.13).

We now show that the perturbations (3.4) are bounded in the sense of Definition 2.1 of bounded perturbation resilience.

Lemma 3.1. *Assume that there exists $q \in \mathbb{R}$ such that $\|p(z^k)\| \leq q$ and $\|p(z^k + \beta_k b^k)\| \leq q$ for all $k \leq K$. Then the perturbations $\beta_k b^k$ defined in (3.4) are bounded for all $k \geq 0$.*

Proof. For all $k \geq 0$, we have $\frac{k-1}{k+2} < 1$. With $\{\beta_k\}_{k=0}^\infty$ chosen as described in (3.4) and $\{\lambda_k^{NE}\}_{k=0}^\infty$ bounded from above by $\Lambda \in \mathbb{R}_{\geq 0}$, it suffices to show that $\|z^k - z^{k-1}\|$ is bounded. If $c(z^{k-1}) = \text{false}$, then

$$z^k = T(z^{k-1}) = z^{k-1} + \lambda(z^{k-1})p(z^{k-1}).$$

Therefore,

$$\|z^k - z^{k-1}\| = \|\lambda(z^{k-1})p(z^{k-1})\| \leq 2q.$$

Otherwise, we get

$$\|z^k - z^{k-1}\| = \|\beta_{k-1} b^{k-1} + \lambda p(z^{k-1} + \beta_{k-1} b^{k-1})\|.$$

We show by induction that this norm is bounded. $k = 2$ is the first iteration index k for which $c(z^{k-1})$ can possibly be true. If $c(z^2) = c(z^1)$ true, then

$$z^3 = T(z^2 + \beta_2 b^2)$$

and

$$\begin{aligned} \|b^2\| &\leq \Lambda \left\| \frac{1}{4} (z^2 - z^1) \right\| \\ &\leq \Lambda \|z^2 - z^1\| \\ &= \Lambda \|\lambda(z^1)p(z^1)\| \\ &\leq \Lambda 2q. \end{aligned}$$

Now, we consider the induction step $k \rightarrow k+1$ and assume that $c(z^{k-1}) = c(z^k)$ true.

$$\begin{aligned}
\|b^k\| &\leq \Lambda \left\| \frac{k-1}{k+2} (z^k - z^{k-1}) \right\| \\
&\leq \Lambda \|z^k - z^{k-1}\| \\
&= \Lambda \|T(z^{k-1} + \beta_{k-1}b^{k-1}) - z^{k-1}\| \\
&= \Lambda \|\beta_{k-1}b^{k-1} + \lambda(z^{k-1} + \beta_{k-1}b^{k-1})p(z^{k-1} + \beta_{k-1}b^{k-1})\| \\
&\leq \Lambda(\beta_{k-1}\|b^{k-1}\| + 2q)
\end{aligned}$$

The induction assumption guarantees that $\|b^{k-1}\|$ are bounded. Therefore

$$\|\beta_{k-1}b^{k-1} + \lambda p(z^{k-1} + \beta_{k-1}b^{k-1})\|$$

is also bounded. For a general sequence $\{\beta_k\}_{k=0}^\infty$ with $\sum_{k=0}^\infty \beta_k < \infty$ and $\beta_k \geq 0$ for all $k \geq 0$, Definition 2.1 calls for a bound on $\|v^k\|$ for all $k \geq 0$. With $\{\beta_k\}_{k=0}^\infty$ chosen as in (2.14), however, it clearly suffices to show a bound on $\|b^k\|$ for $k \leq K$. This completes the proof. \square

The outer perturbation scheme (2.15) using the Nesterov perturbation is

$$y^{k+1} = T(y^k) + \beta_k b^k, \quad (3.5)$$

where

$$b^k = \begin{cases} \lambda_k^{NE} \frac{k-1}{k+2} (y^k - y^{k-1}) - \lambda(y^k)p(y^k), & \text{if } \tilde{c}(y^k, y^{k-1}) = \text{true}, \\ 0, & \text{otherwise,} \end{cases} \quad (3.6)$$

$\{\beta_k\}_{k=0}^\infty$ is chosen as in (2.14), $\{\lambda_k^{NE}\}_{k=0}^\infty$ is a user-chosen bounded sequence of step sizes with $\lambda_k^{NE} \in [0, \Lambda]$ for all $k \geq 0$ and the function \tilde{c} is the same as in (2.12) with c as in (2.13).

Lemma 3.2. *Assume that there exists $q \in \mathbb{R}$ such that $\|p(y^k)\| \leq q$ for all $k \leq K$. Then the perturbations $\beta_k b^k$ defined in (3.6) are bounded for all $k \geq 0$.*

Proof. With $\{\beta_k\}_{k=0}^\infty$ and $\{\lambda_k^{NE}\}_{k=0}^\infty$ chosen as described in (3.5) and bounded from above by $\Lambda \in \mathbb{R}_{\geq 0}$, it suffices to show that

$$\|b^k\| = \Lambda \left\| \frac{k-1}{k+2} (y^k - y^{k-1}) - \lambda(y^k)p(y^k) \right\|$$

is bounded. For all $k \geq 0$, we have $\frac{k-1}{k+2} < 1$ and

$$\begin{aligned}
&\left\| \frac{k-1}{k+2} (y^k - y^{k-1}) - \lambda(y^k)p(y^k) \right\| \\
&\leq \frac{k-1}{k+2} \|y^k - y^{k-1}\| + \lambda(y^k) \|p(y^k)\| \\
&\leq \|y^k - y^{k-1}\| + 2q.
\end{aligned}$$

Therefore we need to show that $\|y^k - y^{k-1}\|$ is bounded. We have $\tilde{c}(y^k, y^{k-1}) = \text{true}$. Hence, we know by construction of \tilde{c} that $\tilde{c}(y^{k-1}, y^{k-2}) = \text{false}$. Therefore,

$$y^k = T(y^{k-1}) = y^{k-1} + \lambda(y^{k-1})p(y^{k-1}).$$

This yields that

$$\|y^k - y^{k-1}\| = \|\lambda(y^{k-1})p(y^{k-1})\| \leq 2q$$

is bounded. \square

The assumptions made in Lemma 3.1 and 3.2, i.e. there exists $q \in \mathbb{R}$ such that $\|p(y^k)\| < q$ or $\|p(z^k)\| < q$ and $\|p(z^k + \beta_k b^k)\| < q$ for all $k \leq K$, is fulfilled if all of the functions φ_i defining the feasible sets of the CFP are convex, which here of course is given by the definition of the CFP. When the functions φ_i are convex for all $i \in I$, the boundedness of $\|p(y^k)\|$, $\|p(z^k)\|$ and $\|p(z^k + \beta_k b^k)\|$ follows from the fact that the norm of the function argument of p is bounded for all $k \leq K$. The latter is the case because K is finite and due to their convexity, the functions φ_i are bounded on any bounded closed subset of \mathbb{R}^n .

4. NUMERICAL DEMONSTRATIONS

In this section, we present numerical results, which illustrate the behaviour of the perturbed level set scheme. We only consider constant relaxation parameter functions and step sizes λ_k^{NE} and therefore we write λ and λ^{NE} instead. The first numerical example is a system of linear inequalities, which is clearly a convex feasibility problem.

4.1. Linear feasibility problem. To demonstrate the main numerical effects via the perturbed and unperturbed cyclic subgradient projection method, we consider a linear feasibility problem, which is stated below. We only present results corresponding to the cyclic projection method because we chose our example in such a way that the simultaneous and cyclic projection method perform very similarly to place our focus on the effect of the Nesterov perturbation.

The linear feasibility problem that we consider is

$$Ax \leq b, \tag{4.1}$$

where

$$A = \delta_{x_2} \begin{pmatrix} \sin(\alpha) \cos(\alpha) & -\sin(\alpha)^2 \\ -\sin(\alpha) \cos(\alpha) & -\sin(\alpha)^2 \end{pmatrix}, b = -\sin(\alpha)^2 \delta_{x_2} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \tag{4.2}$$

for an angle $\alpha \in (0, \pi/2)$ and $\delta_{x_2} \in \mathbb{R}_{\geq 0}$. The feasible half-spaces H_1, H_2 corresponding to the two inequalities given in (4.1) are illustrated in Figure 1. The matrix A is chosen such that the normal vectors v_1, v_2 of the half-spaces H_1, H_2 both include the angle α with the x_1 -axis, i.e.,

$$\langle (-1, 0)^T, v_1 \rangle / \|v_1\| = \langle (1, 0)^T, v_2 \rangle / \|v_2\| = \cos(\alpha).$$

The right hand side b is chosen such that the separating hyperplanes (i.e. lines) of H_1 and H_2 intersect at the point $(0, \delta_{x_2})$. In our calculations we chose $\delta_{x_2} = 10$.

When α is small, the condition number $\kappa(A)$ of course rises as $\langle v_1, v_2 \rangle / (\|v_1\| \|v_2\|) \rightarrow -1$. In this paper, we focus particularly on ill-conditioned problems. So, our focus lies on those cases where the condition number $\kappa(A)$ of the matrix A is big. In Table 1, we give $\kappa(A)$ for $\alpha \in \{5^\circ, 3^\circ, 1^\circ, 0.5^\circ\}$. We observe that $\kappa(A)$ grows significantly when $\alpha \leq 1^\circ$. Therefore the cases $\alpha = 1^\circ$ and $\alpha = 0.5^\circ$ are of particular interest to us.

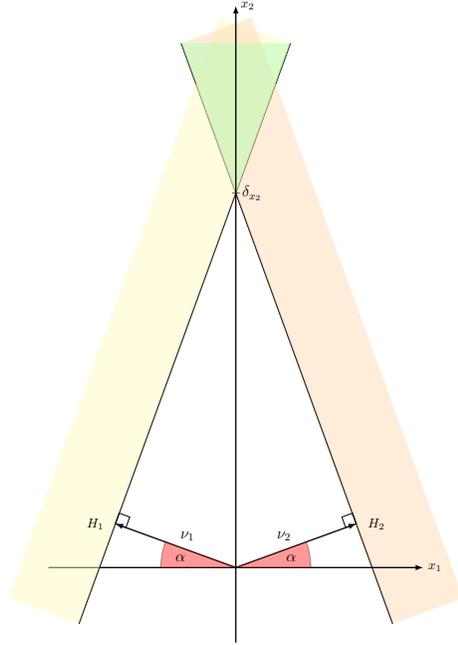


FIGURE 1. Illustration of the system of inequalities (4.1). H_1 and H_2 are the solution sets of the individual inequalities. Their intersection, which is indicated in green, is the set of solutions to (4.1). The normal vectors v_i on H_i ($i = 1, 2$) include the angle α with the x_1 -axis.

TABLE 1. Condition number $\kappa(A)$ of A resulting from different choices of the angle α .

α	5°	3°	1°	0.5°
$\kappa(A)$	11.43	19.08	57.29	114.59

In addition to the angle $\alpha \in \{5^\circ, 3^\circ, 1^\circ, 0.5^\circ\}$, we study the influence of different relaxation parameters λ and consider $\lambda = 1.0, 1.5$ and 1.9 .

In what follows, we occasionally use abbreviations for the cyclic projection method (CP) as well as for the Nesterov perturbation (NE). By K_{CP} , we denote the number of iterations it takes the unperturbed cyclic subgradient projection method to find a solution of (4.1). $K_{\text{CP, NE}}$ denotes the number of iterations required by the perturbed cyclic subgradient projection method to find such a solution (i.e. to terminate). In the following, we consider the ratio $\sigma_{\text{CP, NE}}$, which is defined as

$$\sigma_{\text{CP, NE}} = \frac{K_{\text{CP, NE}}}{K_{\text{CP}}}. \quad (4.3)$$

Values of $\sigma_{\text{CP, NE}} < 1$ serve as an indicator of the acceleration provided by the perturbed cyclic projection compared to its unperturbed counterpart. For our calculations, we choose the starting point $x^0 = (1, 0)^T$ and the parameters $\varepsilon_{\text{max}} = 6 \cdot 10^{-2}$, $\varepsilon_{\text{min}} = 0$. We consider an iterate x^* to be a solution of (4.1) if $\|Ax^* - b\|_\infty \leq 10^{-10}$.

TABLE 2. Values of $\sigma_{\text{CP,NE}}$ for all four values of α and all three values of λ . We refer to values of $K_{\text{CP,NE}}$ resulting from step sizes λ^{NE} , which were chosen experimentally such that $K_{\text{CP,NE}}$ is as small as possible. The perturbed CP terminates significantly faster than the unperturbed CP in most cases. Only $\lambda = 1.9$ and $\alpha = 5^\circ$ or 3° result in $\sigma_{\text{CP,NE}} > 1$.

λ	$\sigma_{\text{CP,NE}}$			
	$\alpha = 5^\circ$	$\alpha = 3^\circ$	$\alpha = 1^\circ$	$\alpha = 0.5^\circ$
1.0	$3.26 \cdot 10^{-3}$	$1.22 \cdot 10^{-3}$	$1.51 \cdot 10^{-4}$	$4.04 \cdot 10^{-5}$
1.5	$4.86 \cdot 10^{-2}$	$3.28 \cdot 10^{-2}$	$1.24 \cdot 10^{-2}$	$6.74 \cdot 10^{-3}$
1.9	1.42	1.29	$8.01 \cdot 10^{-2}$	$3.92 \cdot 10^{-2}$

In Table 2, we present the values of $\sigma_{\text{CP,NE}}$ for $\alpha \in \{5^\circ, 3^\circ, 1^\circ, 0.5^\circ\}$ and $\lambda \in \{1.0, 1.5, 1.9\}$. The values of $K_{\text{CP,NE}}$ we refer to result from step sizes λ^{NE} , which are chosen experimentally such that $K_{\text{CP,NE}}$ is as small as possible. The values in Table 2 indicate that the perturbed cyclic projection method terminates significantly faster than its unperturbed counterpart, especially when α is small and therefore $\kappa(A)$ is big. However, there are two instances when $\sigma_{\text{CP,NE}} > 1$. Those occur when $\lambda = 1.9$ and $\alpha = 5^\circ$ or 3° . Note that in these only slightly ill-conditioned cases the overrelaxed unperturbed projection method terminates faster than its perturbed counterpart.

Figure 2 illustrates how K_{CP} and $K_{\text{CP,NE}}$ increase for decreasing α and $\lambda \in \{1.0, 1.5, 1.9\}$. We observe that K_{CP} increases by at least one order of magnitude for $\alpha \leq 1$ while the increase of $K_{\text{CP,NE}}$ is far more moderate and its values for $\alpha \leq 1^\circ$ are at least one order of magnitude lower than those of K_{CP} for all $\lambda \in \{1.0, 1.5, 1.9\}$.

The trajectory of iterates presented in Figure 3(a) corresponds to $\alpha = 5^\circ$ and $\lambda = 1.5$. It illustrates why the aforementioned increase of K_{CP} is to be expected for small α . As α gets smaller, the zigzagging gets more severe and can not even be compensated by the beneficial effect of overrelaxation, which amplifies the small progress in x_2 -direction contained in each projection step $p(x^k)$.

In Figure 3(b), we demonstrate how the Nesterov perturbation can overcome these difficulties. Again, we consider $\alpha = 5^\circ$ and $\lambda = 1.5$ and depict the trajectory of the perturbed iteration by using the step size $\lambda^{NE} = 2$. The progress in x_2 -direction is here increased by the perturbation steps. Furthermore, an iterate y^k resulting from perturbed steps violate one of the inequalities in a more severe manner than an iterate x^k that we would get from an unperturbed iteration. Therefore the subsequent projection step $p(y^k)$ is large, which again amplifies the movement in x_2 -direction. This effect is increased even more by a relaxation parameter $\lambda > 1$.

However, the same mechanism can lead to numerical difficulties when $\lambda > 1$ and we choose λ^{NE} too big. Then the more severe violation of the inequalities combined with the overrelaxation leads the iterates to regions where computers fail to evaluate Ax correctly while in theory the iterates still converge to a solution of (4.1) if $K \in \mathbb{N}$. If any knowledge on the magnitude of the solutions to (4.1) exists it is therefore a good idea in order to circumvent that behavior to e.g. only perturb an iteration if the coordinates of x^k are within certain intervals and continue the iteration in an unperturbed manner otherwise.

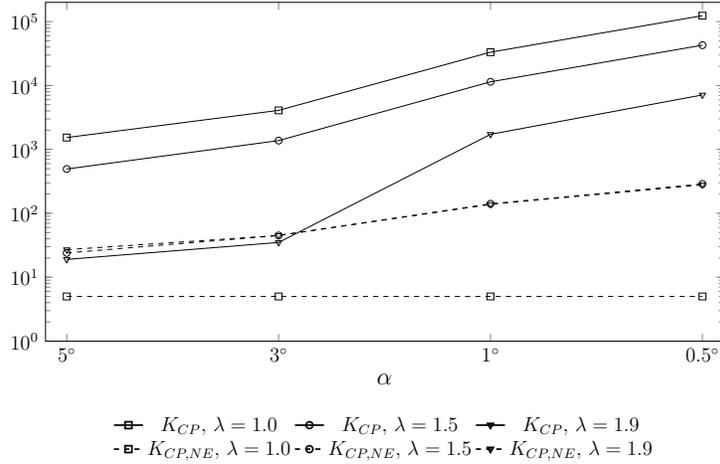
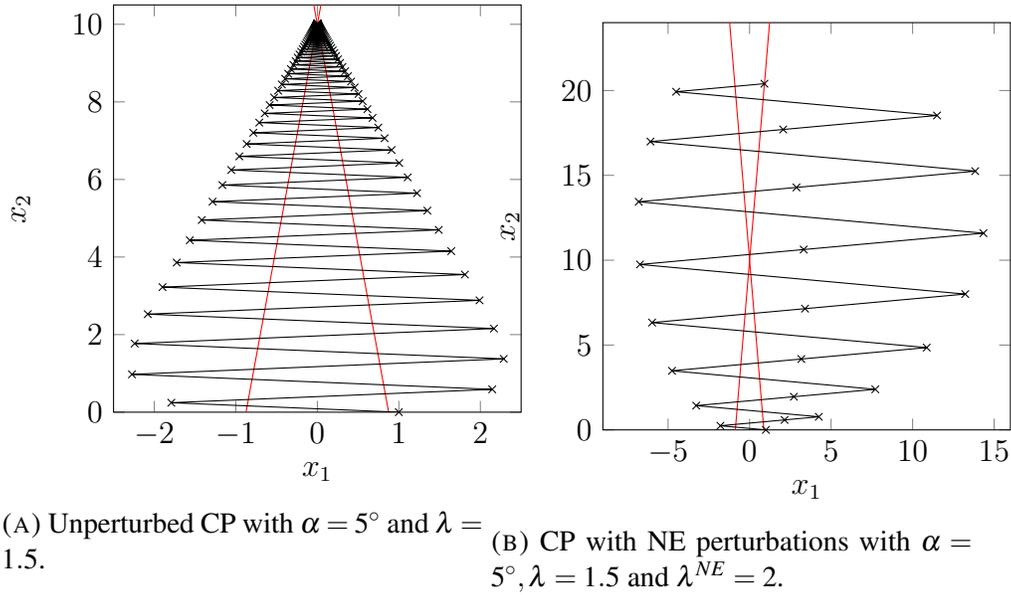


FIGURE 2. K_{CP} and $K_{CP,NE}$ increase for decreasing α for all $\lambda \in \{1.0, 1.5, 1.9\}$. K_{CP} increases by at least one order of magnitude for $\alpha \leq 1$ while the increase of $K_{CP,NE}$ is far more moderate and its values for $\alpha \leq 1^\circ$ are at least one order of magnitude lower than those of K_{CP} for all $\lambda \in \{1.0, 1.5, 1.9\}$



(A) Unperturbed CP with $\alpha = 5^\circ$ and $\lambda = 1.5$. (B) CP with NE perturbations with $\alpha = 5^\circ$, $\lambda = 1.5$ and $\lambda^{NE} = 2$.

FIGURE 3. Trajectories of the iterates resulting from perturbed and unperturbed CP. \times marks the individual iterates. The separating hyperplanes of H_1 and H_2 are drawn in red.

The aforementioned step sizes λ^{NE} that we use to produce the results presented in Table 2 are given in Table 3. We observe that, for a relaxation parameter $\lambda = 1.5$ or $\lambda = 1.9$, the chosen step sizes λ^{NE} do not vary much, but for $\lambda = 1$, they increase drastically. This happens due to the fact that in this simple linear example the best strategy for $\lambda = 1.0$ is to choose λ^{NE} so big that the x_2 -coordinate of the first perturbed iterate y^k is $\geq \delta_{x_2}$. Thereby y^{k+1} is a solution to (4.1), because it is an unperturbed iterate and as $\lambda = 1.0$ it is an exact (in contrast to overrelaxed)

TABLE 3. Step sizes.

λ	λ^{NE}			
	$\alpha = 5^\circ$	$\alpha = 3^\circ$	$\alpha = 1^\circ$	$\alpha = 0.5^\circ$
1.0	90	230	2500	9000
1.5	2	2	1.6	1.55
1.9	0.5	0.25	0.25	0.2

metric projection onto the set of solutions to (4.1). Employing this strategy we get a constant number $K_{CP, NE}$ of 5 iterations for any value of α , as shown in Figure 2.

4.2. Intensity-modulated radiation therapy. Intensity-modulated radiation therapy (IMRT) is used as one treatment option in clinical oncology. With it, physicians aim to destroy the tumor tissue with irradiation while sparing surrounding healthy organs as much as possible. These goals are conflicting because of physical limitations like the proximity of the tumor to healthy organs and the attainable decrease of the dose values over a given distance. Therefore, the treatment planner usually has to decide for a compromise between several goals.

The IMRT planning problem, as we consider it in this paper, is to determine optimal intensity maps, i.e. a set of fluence intensity values, which causes a dose distribution in the patient's body which will best fulfill the clinical goals formulated by the physician. In contrast to the previous examples, the IMRT planning problem is a nonlinear one. The vector d of dose values received by each voxel in the patient's body when the fluence intensity given by the vector x is applied can be calculated using the so-called dose matrix P , which only contains non-negative values. Due to physical reasons, the vector x is restricted to be non-negative, which results in dose values $d = P \cdot x$, which are also non-negative.

We use the following set of functions to represent the dose prescriptions given by the clinical goals. All functions refer to a biological structure and evaluate the dose received by the voxels contained in the structure. From a mathematical point of view that structure is a set \mathcal{O} of indices which correspond to the voxels contained in the biological structure. By $d_i = \langle p_i, x \rangle$ we denote the dose value received by the voxel with index i , where p_i is the i -th row of P .

The *quadratic upper tail penalty function* penalizes dose values of d corresponding to the structure \mathcal{O} which exceed a given threshold $U \in \mathbb{R}$

$$f(d, \mathcal{O}) = \frac{1}{|\mathcal{O}|} \sum_{i \in \mathcal{O}} \max(0, d_i - U)^2.$$

The *quadratic lower tail penalty function* penalizes dose values of d corresponding to the structure \mathcal{O} which fall below a given threshold $L \in \mathbb{R}$

$$f(d, \mathcal{O}) = \frac{1}{|\mathcal{O}|} \sum_{i \in \mathcal{O}} \max(0, L - d_i)^2.$$

The *Equivalent uniform dose (EUD) function* is related to the well-known EUD concept of Niemierko [28]. It penalizes dose values of d corresponding to the structure \mathcal{O} which deviate from 0

$$f(d, \mathcal{O}) = \frac{1}{|\mathcal{O}|} \sum_{i \in \mathcal{O}} d_i^q, \quad q \in (1, \infty).$$

TABLE 4. Dose evaluation functions of the considered biological structures with exemplary (i.e. corresponding to one of the IMRT cases) number of voxels $|\mathcal{O}|$ belonging to the respective biological structures. The dimension of the optimization variable x is 1725 for the IMRT case we are referring to.

	structure	function name	parameters	$ \mathcal{O} $
f_1	Left parotis	EUD	$q = 2$	95
f_2	Right parotis	EUD	$q = 2$	103
f_3	Myelon	EUD	$q = 2$	81
f_4	Unclassified healthy tissue	EUD	$q = 2$	42118
f_5	Tumor volume	Tumor conformity	$q = 2, d^{ref} = 60$	4899
g_1	Tumor volume	lower tail penalty	$L = 55$	4899
g_2	Tumor volume	upper tail penalty	$U = 66$	4899
g_3	Myelon	upper tail penalty	$U = 45$	81

The *tumor conformity function* is used to ensure an even dose distribution within the tumor volume. It penalizes dose values of d corresponding to the structure \mathcal{O} which deviate from a given reference value

$$f(d, \mathcal{O}, d^{ref}) = \frac{1}{|\mathcal{O}|} \sum_{i \in \mathcal{O}} |d^{ref} - d_i|^q \quad q \in (1, \infty).$$

All of these functions take into account the system of linear (in)equalities resulting from $d_i = \langle p_i, x \rangle$ being less/ greater or equal than a right hand side value determined by the function parameters, where $i \in \mathcal{O}$. The functions measure the distance of the vector x to the (separating) hyperplanes corresponding to the linear (in)equalities in a nonlinear way. The given functions do not represent a complete list of dose evaluation functions used in clinical applications. A more complete survey of such functions can be found in [33] and [35].

4.2.1. The IMRT optimization problem. In this paper, we choose four head neck cases for our numerical experiments. We consider a reduced set of biological structures and their dose evaluation functions, which focuses on the main conflict between irradiating the tumor volume on the one hand and sparing the myelon and parotids on the other hand. Additionally, we include the healthy tissue not associated with any of the structures mentioned above with the goal to keep the dose in this tissue as low as possible. The dose evaluation functions we use are given in Table 4 in detail.

We use a reduced set of biological structures for our calculations, because our focus is to demonstrate the effects of the mathematical methods. Therefore, the treatment plans resulting from our calculations do not fulfill all of the clinical goals a treatment planner would formulate for a full head neck case.

The IMRT optimization problem as we formulate it in this work is

$$\text{Minimize } \Phi(x) \quad (4.4)$$

$$\text{s.t. } g_j(x) \leq 0, \quad j \in J, \quad (4.5)$$

where $\Phi = \sum_{i \in I} w_i f_i$, $I = \{1, 2, 3, 4, 5\}$, $J = \{1, 2, 3, 4\}$, $w_i \in \mathbb{R}_{\geq 0}$ with $\sum_{i \in I} w_i = 1$ and f_i, g_1, g_2, g_3 as listed in Table 4. All of the functions given in Table 4 operate on $d = Px$, while the function g_4 , which ensures the positivity of the fluence intensity vector $x \in \mathbb{R}^n$, is chosen as

$$g_4(x) = \frac{1}{n} \sum_{s=1}^n \min(x_s, 0)^2$$

and operates on $x \in \mathbb{R}^n$, where x_s are the coordinates of x .

4.2.2. Results. We stop the algorithm if the projection method we use is unable to find a feasible solution of the current CFP after $n_{\max} = 10^3$ iterations. Then, the current CFP is assumed to be infeasible and we consider the solution of the previous CFP to be the result of the perturbed level set scheme. In all our calculations we choose $\varepsilon_{\min} = 0$ and $\lambda = 1.9$.

In what follows, we first compare results on the Nesterov perturbation for different threshold values ε_{\max} , next we keep ε_{\max} fixed and illustrate the results of choosing different step sizes λ^{NE} and finally we compare the results produced by the Nesterov perturbation to those achieved by the heavy ball and surrogate constraint perturbation, which were introduced in [3]. By Φ_{method}^* , we denote the lowest objective function value, for which the level set scheme using the specified unperturbed method is able to find a solution within the given maximum number n_{\max} of iterations per CFP P^s as described in (2.4). The methods we consider are the cyclic (CP) and simultaneous projection (SP) method. In a similar manner $\Phi_{\text{method, pert}}^*$ denotes the lowest objective function value achieved by the perturbed level set scheme using the specified method and perturbation. The perturbations we consider in the following are the Nesterov (NE), the heavy ball (HB) and the surrogate constraint (SC) perturbation. By K_{method} , we denote the total number of iterations it takes the level set scheme using the specified unperturbed method to find the solution x_{method}^* with the objective function value $\Phi_{\text{method}}^* = \Phi(x_{\text{method}}^*)$. $K_{\text{method, pert}}$ denotes the total number of iterations required by the perturbed level set scheme using the specified method and perturbation to find the solution $x_{\text{method, pert}}^*$ with the objective function value $\Phi_{\text{method, pert}}^* = \Phi(x_{\text{method, pert}}^*)$. Furthermore we denote by $\hat{K}_{\text{method, pert}}$ the total number of iterations required by the perturbed level set scheme using the specified method and perturbation to find an iterate solution $\hat{x}_{\text{method, pert}}$ with a smaller objective function value than the minimal value that was found using the unperturbed method, i.e. $\Phi(\hat{x}_{\text{method, pert}}) \leq \Phi_{\text{method}}^*$.

We consider the ratios $\sigma_{\text{method, NE}}$, $\hat{\sigma}_{\text{method, NE}}$ and $\theta_{\text{method, NE}}$, which are defined as follows:

$$\sigma_{\text{method, pert}} = \frac{K_{\text{method, pert}}}{K_{\text{method}}}, \quad (4.6)$$

$$\hat{\sigma}_{\text{method, pert}} = \frac{\hat{K}_{\text{method, pert}}}{K_{\text{method}}}, \quad (4.7)$$

$$\theta_{\text{method, pert}} = \frac{\Phi_{\text{method, pert}}^*}{\Phi_{\text{method}}^*}. \quad (4.8)$$

TABLE 5. K_{method} and Φ_{method}^* for (unperturbed) SP and CP. $K_{\text{SP}} < K_{\text{CP}}$ for two and $\Phi_{\text{SP}}^* < \Phi_{\text{CP}}^*$ for three out of four IMRT cases.

	K_{SP}	K_{CP}	Φ_{SP}^*	Φ_{CP}^*
case 1	7159	6665	3480	3563
case 2	3523	4484	2378	2424
case 3	4773	4238	3129	3178
case 4	4496	5280	3098	3093

All three ratios provide a measure of improvement or acceleration with respect to the unperturbed projection methods. Values significantly smaller than 1 correspond to big improvements or accelerations.

4.2.3. *Varying ε_{max} .* For our first set of results, we choose $\lambda^{NE} = 1$ and varied ε_{max} . We present results corresponding to three different choices $\varepsilon_{\text{max}}^1, \varepsilon_{\text{max}}^2$ and $\varepsilon_{\text{max}}^3$ of ε_{max} such that

$$\begin{aligned} \cos^{-1}(-1 + \varepsilon_{\text{max}}^1) &= 165^\circ, \\ \cos^{-1}(-1 + \varepsilon_{\text{max}}^2) &= 150^\circ, \\ \text{and } \cos^{-1}(-1 + \varepsilon_{\text{max}}^3) &= 135^\circ. \end{aligned}$$

In Figures 4 and 5, we illustrate how the three choices of ε_{max} influence the values of $\sigma_{\text{method, NE}}$ and $\theta_{\text{method, NE}}$ for all four IMRT cases. Figure 4 refers to the perturbed level set scheme using the simultaneous projection method in combination with the Nesterov perturbation while the values presented in Figure 5 result from using the cyclic projection method, again in combination with the Nesterov perturbation.

In Figure 4, we see that the perturbed simultaneous projection method converges significantly faster than its unperturbed counterpart, which results in values of $\sigma_{\text{SP, NE}}$ between 0.2137 and 0.5961 with a mean of 0.4403 (all values rounded to 10^{-4}). However, for IMRT case 1 only $\varepsilon_{\text{max}}^3$ leads to $\theta_{\text{SP, NE}} < 1$, i.e. $\Phi_{\text{SP, NE}}^* < \Phi_{\text{SP}}^*$.

Figure 5 shows that for two out of the four IMRT cases $\varepsilon_{\text{max}}^1$ lead to a situation where the condition for perturbing an iteration was never met. The values of $\sigma_{\text{CP, NE}}$ range from 0.3947 to 1.2020 with a mean of 0.6605.

Among the three variants $\varepsilon_{\text{max}}^1, \varepsilon_{\text{max}}^2$ and $\varepsilon_{\text{max}}^3$ presented here, $\varepsilon_{\text{max}}^2$ is the most favorable choice with respect to two aspects. First, in contrast to $\varepsilon_{\text{max}}^1$ it ensures that the condition for applying perturbations is met at some point during the iteration process for all four IMRT cases even when we use the cyclic projection method. Second, for the majority of instances considered it results in fewer iterations required compared to $\varepsilon_{\text{max}}^3$.

In Figure 6, we illustrate in an exemplary manner for IMRT case 1 when and how often an iteration is perturbed. As is to be expected the perturbations occur earlier during the iteration process for the threshold value $\varepsilon_{\text{max}}^3$ than for $\varepsilon_{\text{max}}^1$, because it is a less strict criterion. We furthermore observe that especially for a stricter threshold value like $\varepsilon_{\text{max}}^1$, but also for $\varepsilon_{\text{max}}^2$ the perturbed level set scheme using the cyclic projection method the condition for perturbing a given iteration is fulfilled later than when we use the simultaneous projection method. This phenomenon occurs due to the different mechanics of the two projection methods. Subsequent projection steps of the cyclic projection method include only gradient information of

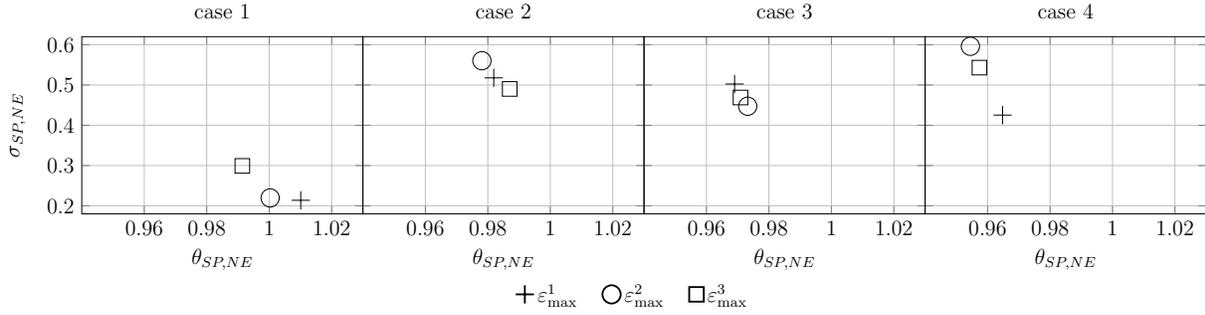


FIGURE 4. Values of $\sigma_{SP,NE}$ plotted against $\theta_{SP,NE}$ resulting from IMRT optimization for different threshold values ϵ_{\max} .

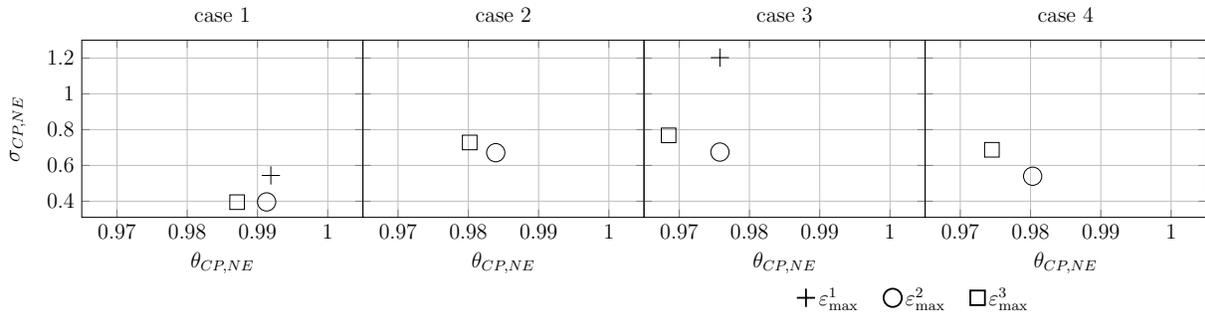


FIGURE 5. Values of $\sigma_{CP,NE}$ plotted against $\theta_{CP,NE}$ resulting from IMRT optimization for different threshold values ϵ_{\max} .

functions defining those (violated) constraints which are subsequent with respect to the control sequence. The simultaneous projection method on the other hand includes gradient information from all violated constraints in every projection step. When applied to an optimization problem with strongly opposing goals like IMRT often two groups of constraints emerge, which are fulfilled/ violated in every other iteration. The projection steps in subsequent iterations then are weighted sums of function gradients where the functions are contained in said opposing groups corresponding to opposing goals. Therefore subsequent projection steps of the simultaneous projection method tend to fulfill the condition which determines whether an iteration is to be perturbed much earlier during the iteration process than those of the cyclic projection method. There this inherent conflict of multiple functions is only revealed later during the iteration process when subsequent (with respect to the control sequence) violated constraints happen to be defined by functions with opposing gradients.

In addition to these exemplary illustrations, we give the number of total perturbed iterations used during the perturbed level set scheme in Table 6. When the cyclic projection method is used to solve the arising CFPs the condition for perturbing a given iteration is not met in any iteration for two of the four IMRT cases when we chose the threshold value ϵ_{\max}^1 . For 6 among the remaining 10 instances, fewer iterations are perturbed when the simultaneous projection method is used.

In Table 7, we present the fractions $\hat{\sigma}_{SP,NE}$ and $\hat{\sigma}_{CP,NE}$ for all four IMRT cases and the three considered options for ϵ_{\max} (values rounded to 10^{-4}). For IMRT case 1 the threshold values ϵ_{\max}^1 and ϵ_{\max}^2 used in combination with the perturbed simultaneous projection method lead to

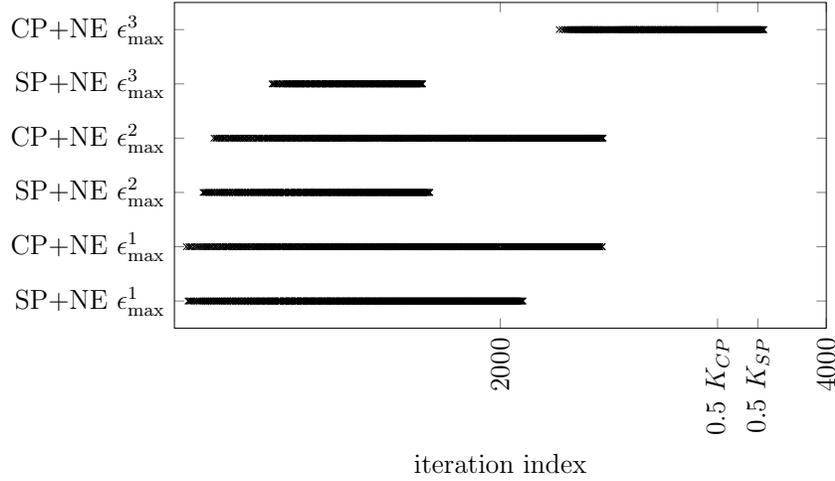


FIGURE 6. Indices of perturbed iterations for perturbed SP and CP using different threshold values ϵ_{\max} in an exemplary manner for IMRT case 1. CP is perturbed later during the iteration for ϵ_{\max}^2 and significantly later for ϵ_{\max}^3 .

TABLE 6. The number of perturbed iterations occurring in the perturbed level set scheme using the simultaneous and cyclic projection method. When we use CP together with the threshold parameter ϵ_{\max}^1 the condition for perturbing an iteration is never met for two of four IMRT cases. Out of the other 10 instances, using SP results in fewer perturbed iterations than using CP for 6 instances.

	SP + NE			CP + NE		
	ϵ_{\max}^1	ϵ_{\max}^2	ϵ_{\max}^3	ϵ_{\max}^1	ϵ_{\max}^2	ϵ_{\max}^3
case 1	263	382	582	275	479	489
case 2	349	539	480	-	607	703
case 3	558	610	656	266	583	703
case 4	455	789	722	-	582	808

$\Phi_{\text{SP,NE}}^* > \Phi_{\text{SP}}^*$. Therefore no value for $\hat{\sigma}_{\text{SP,NE}}$ exists. For two out of four IMRT cases the condition triggering the perturbation of an iteration was never met when using the cyclic projection method in combination with the threshold value ϵ_{\max}^1 . For these instances, we have $\hat{\sigma}_{\text{CP,NE}} = 1$. For all other instances $\hat{\sigma}_{\text{SP,NE}}$ is significantly smaller than $\hat{\sigma}_{\text{CP,NE}}$, i.e. the perturbed simultaneous projection method arrives at Φ_{method}^* faster than the perturbed cyclic projection method. The values of $\hat{\sigma}_{\text{SP,NE}}$ range from 0.2313 to 0.3684 with a mean of 0.2880 while $\hat{\sigma}_{\text{CP,NE}}$ ranges from 0.3961 to 1 with a mean of 0.6182.

TABLE 7. The fractions $\hat{\sigma}_{\text{SP,NE}}$ and $\hat{\sigma}_{\text{CP,NE}}$ for all four IMRT cases and the three considered options for ε_{\max} . For IMRT case 1 the threshold values ε_{\max}^1 and ε_{\max}^2 used in combination with the perturbed SP lead to $\Phi_{\text{SP,NE}}^* > \Phi_{\text{SP}}^*$. Therefore no value for $\hat{\sigma}_{\text{SP,NE}}$ exists. For all other instances $\hat{\sigma}_{\text{SP,NE}}$ is significantly smaller than $\hat{\sigma}_{\text{CP,NE}}$, i.e. the perturbed SP arrives at Φ_{method}^* faster than the perturbed CP.

	ε_{\max}^1	$\hat{\sigma}_{\text{SP,NE}}$ ε_{\max}^2	ε_{\max}^3	ε_{\max}^1	$\hat{\sigma}_{\text{CP,NE}}$ ε_{\max}^2	ε_{\max}^3
case 1	-	-	0.2982	0.5439	0.3961	0.3928
case 2	0.3684	0.3398	0.3295	1	0.5256	0.5230
case 3	0.2690	0.2747	0.2686	0.9339	0.4809	0.5621
case 4	0.2578	0.2313	0.2429	1	0.5384	0.5220

Next, we illustrate the effect of varying the Nesterov perturbation step size λ^{NE} while keeping ε_{\max} fixed. As discussed before, ε_{\max}^2 is the most favorable choice among the values we considered so we choose ε_{\max} such that $\cos^{-1}(-1 + \varepsilon_{\max}) = 150^\circ$ in the following.

The values of the Nesterov perturbation step size λ^{NE} we consider here are 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75 and 2. The values of the ratios $\theta_{\text{method,NE}}$ and $\sigma_{\text{method,NE}}$ resulting from these values are presented in Figure 7. Bigger perturbation steps accelerate the iteration process more than smaller ones, but also make it harder for the perturbed level set scheme to stay below $\Phi_{\text{method,NE}}^*$ resulting in values of $\theta_{\text{method,NE}}$ close to or even bigger than 1. Using cyclic projection in combination with the Nesterov perturbation and a step size of 0.5 or smaller results for three out of four IMRT cases in values of $\sigma_{\text{CP,NE}} > 1$, which originates from slower termination than the unperturbed cyclic projection method. The perturbed simultaneous projection method on the other hand achieves values $\sigma_{\text{SP,NE}} \leq 1$ for all step size choices, i.e. it terminates at least as fast as the unperturbed method.

4.2.4. Comparison with HB and SC perturbation. In this section, we compare the results achieved by the Nesterov perturbation to those produced by the heavy ball and surrogate constraint perturbation, which were introduced in [3]. In contrast to the Nesterov perturbation, the heavy ball and surrogate constraint perturbation construct a new direction from the gradient information at the current and previous iterate. Here we only consider the simultaneous projection as basic method to solve the CFPs $\{P^s\}$ because it has performed better in previous experiments by employing perturbations earlier in the iteration process.

We present results achieved by two sets of step sizes: The first set contains naive step sizes while for the second set the step sizes are chosen in an experimental way such that the perturbed level set scheme achieves particularly low objective function values.

To quantify the acceleration achieved by the perturbed level set scheme for all four IMRT cases we illustrate in Figures ...and 9 how the methods compare with respect to the values of $\theta_{\text{SP,pert}}$ and $\sigma_{\text{SP,pert}}$ they are able to achieve. Both sets of step sizes cause the perturbed iteration scheme to arrive at solutions with a lower objective function value within fewer iterations. Using naive step sizes $\lambda^{NE} = 1, \lambda^{SC} = 1, \lambda^{HB} = 1$, the Nesterov perturbation accelerates the perturbed level set scheme more than the heavy ball and less than the surrogate constraint perturbation.

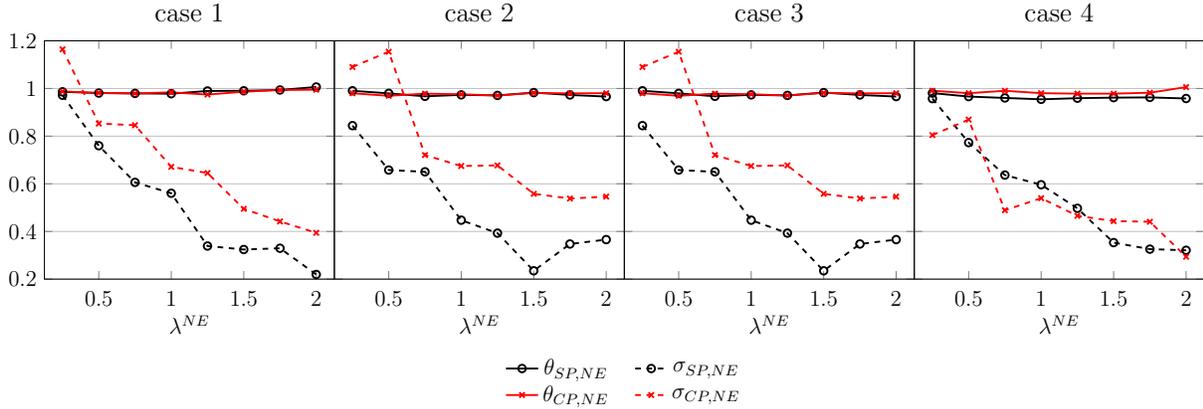


FIGURE 7. Values of $\theta_{\text{method, NE}}$ and $\sigma_{\text{method, NE}}$ resulting from IMRT optimization problems for different choices of the Nesterov perturbation step size λ^{NE} . Bigger step sizes tend to result in lower values of $\sigma_{\text{method, NE}}$, i.e. faster termination of the algorithm. However, not all values of $\theta_{\text{method, NE}}$ for big step sizes are ≤ 1 , which means that in these instances we have $\Phi_{\text{method, NE}}^* > \Phi_{\text{method}}^*$. Perturbation step sizes smaller than 0.5 in combination with the cyclic projection method tend to lead to values of $\sigma_{CP, NE} > 1$, which means that the termination of the algorithm is slowed down with respect to the unperturbed cyclic projection method. All choices of λ^{NE} lead to $\sigma_{SP, NE} \leq 1$, i.e. the perturbed simultaneous projection method terminates at least as fast as its unperturbed counterpart.

The second set of results shows step sizes λ^{NE} , λ^{HB} and λ^{SC} tuned individually for each IMRT case offer advantages in terms of iteration numbers over naive step sizes, especially for the heavy ball perturbation. Still all perturbed schemes achieve lower objective function values within fewer iterations than the unperturbed scheme. When the step sizes are chosen such that particularly low objective function values are achieved, the heavy ball and surrogate constraint perturbation accelerate the perturbed level set scheme even more than the Nesterov perturbation. In practice, however, tuning step sizes for each instance is impracticable and either naive step sizes or step sizes which in prior experiments have shown to work well for a wide number of similar instances would be used. Our results imply that even though for Nesterov perturbation we do not need to construct a new, non-zigzagging direction, it is in such situations a reliable option to accelerate the unperturbed projection method.

5. CONCLUSIONS

In this paper, we demonstrate how the level set can be used to transform a general convex constrained optimization problem into a sequence of convex feasibility problems. We present both the cyclic and simultaneous subgradient projection method as tools to solve the arising feasibility problems. Furthermore, we exploit the fact that both projection methods are bounded perturbation resilient and modified their iteration scheme. To this aim we introduce the Nesterov perturbation. Solving the feasibility problems arising from the level set scheme via perturbed projection methods is what we call the perturbed level set scheme. In our numerical results we examine its behavior, in particular for ill-conditioned problems.

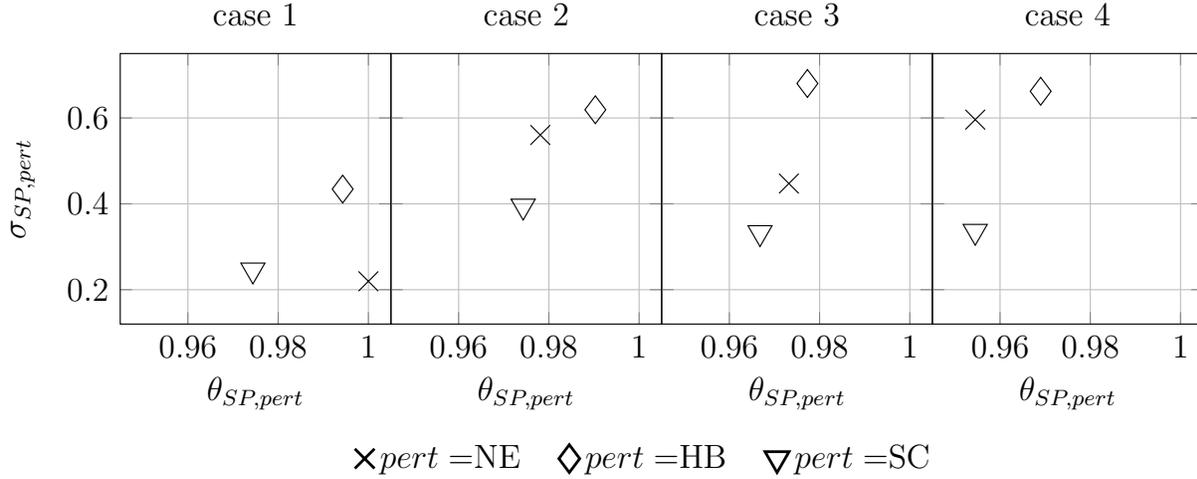


FIGURE 8. $\theta_{SP, pert}$ plotted against $\sigma_{SP, pert}$ achieved by the NE, HB and SC perturbation using naive step sizes. All perturbed schemes achieve lower or equal objective function values than the unperturbed one within fewer iterations. The NE and SC perturbation lead to termination of the algorithm than the HB perturbation.

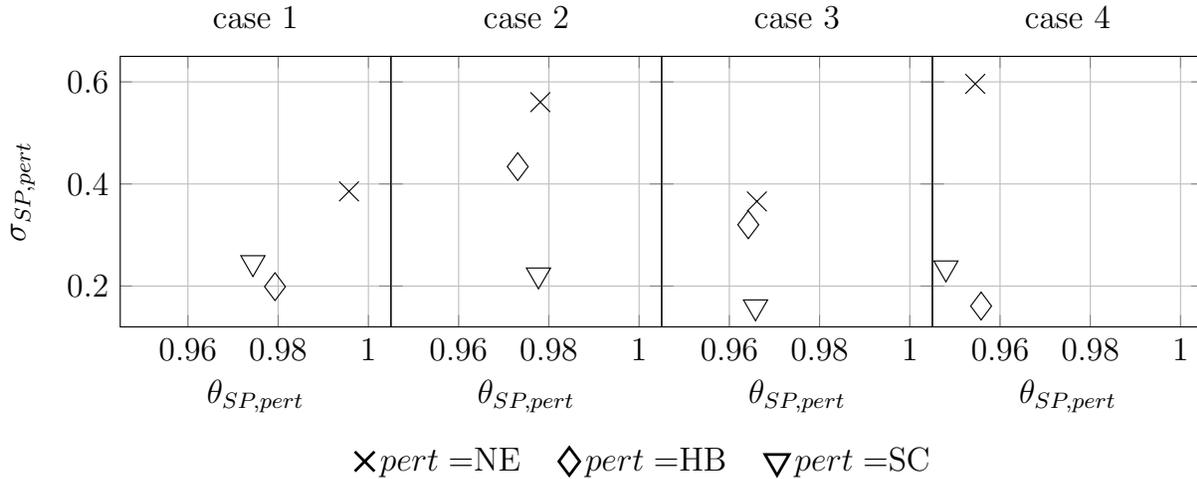


FIGURE 9. $\theta_{SP, pert}$ plotted against $\sigma_{SP, pert}$ achieved by the NE, HB and SC perturbation. All perturbed schemes achieve lower objective function values than the unperturbed one within fewer iterations. The HB and SC perturbation lead to termination of the algorithm than the NE perturbation.

We present numerical results on a linear system of inequalities, which illustrate the mechanism of the Nesterov perturbation. We observe that in the majority of the considered instances the use of the perturbed projection methods result in significantly faster termination of the algorithm, especially when the system is ill-conditioned. Next we apply the perturbed level set scheme to IMRT optimization problems arising from four IMRT head neck cases. We vary the threshold parameter ε_{\max} and observe that for all considered choices of ε_{\max} the perturbed level set scheme using the simultaneous projection and Nesterov perturbation terminates significantly

faster than its unperturbed counterpart. The mean percentage of iteration numbers required by the perturbed level set scheme using the simultaneous projection method compared to its unperturbed version is 44%. When we use the cyclic projection method, the condition to trigger the perturbation of an iteration is not triggered for all IMRT cases and choices of ε_{\max} . The mean percentage of iteration numbers required by the perturbed level set scheme using the cyclic projection method compared to its unperturbed version is 66%. Next, we keep ε_{\max} fixed and vary the perturbation step size λ^{NE} . Bigger step sizes tend to result in lower iteration numbers, but make it slightly harder for the algorithm to produce a solution, which is at least as good as the one resulting from the unperturbed method measured in terms of the objective function.

Finally we compare the results achieved using the Nesterov perturbation to those produced using the heavy ball and surrogate constraint perturbation introduced in [3]. All perturbations achieve solutions with lower objective function values within fewer iterations than the unperturbed level set scheme. When the step sizes were chosen by experiment individually for each IMRT case, The Nesterov perturbation offered less acceleration of the algorithms termination than the heavy ball and surrogate constraint perturbation. However, when the step sizes were chosen with less knowledge from a priori experiments (which would be the most likely scenario for an actual application), the Nesterov perturbation achieved better results than the heavy ball perturbation and in some of the IMRT cases came close to the performance of the surrogate constraint perturbation.

Our observations suggest that the Nesterov perturbation can be used as a tool to accelerate the termination of the algorithm when zigzagging behavior occurs in ill-conditioned problems. Even though for the Nesterov perturbation no new non-zigzagging directions are constructed, as it is done for the heavy ball and surrogate constraint perturbation, its performance is in some situations of practical relevance comparable to or even better than those of the heavy ball and surrogate constraint perturbation.

Acknowledgements

We thank the referees for their thorough analysis and review, all their comments and suggestions helped tremendously in improving the quality of this paper.

REFERENCES

- [1] D.P. Bertsekas, *Nonlinear Programming*, (2nd edition), Athena Scientific, Belmont, MA, 1999.
- [2] C. Bargetz, S. Reich, R. Zalas, Convergence properties of dynamic string averaging projection methods in the presence of perturbations, *Numer. Algo.* 77 (2017), 185-209.
- [3] E. Bonacker, A. Gibali, K.-H. Küfer, Accelerating two projection methods via perturbations with application to intensity-modulated radiation therapy, *Appl. Math. Optim.* (2019), doi: 10.1007/s00245-019-09571-4.
- [4] E. Bonacker, A. Gibali, K.-H. Küfer, P. Süß, Speedup of lexicographic optimization by superiorization and its applications to cancer radiotherapy treatment, *Inverse Probl.* 33 (2017), 044012.
- [5] Y. Censor, Superiorization and perturbationresilience of algorithms: A bibliography compiled and continuously updated. <https://arxiv.org/abs/1506.04219>. Online at: <http://math.haifa.ac.il/yair/bib-superiorization-censor.html> (last updated: March 25, 2020.)
- [6] Y. Censor, Weak and strong superiorization: Between feasibility-seeking and minimization, *Analele Stiintifice ale Universitatii Ovidius Constanta-Seria Matematica* 23 (2015), 41-54.
- [7] Y. Censor, Can linear superiorization be useful for linear optimization problems? *Inverse Prob.* 33 (2017), 044006.

- [8] Y. Censor, W. Chen, H. Pajoohesh, Finite convergence of a subgradient projections method with expanding controls, *Appl. Math. Optim.* 64 (2011), 273–285.
- [9] Y. Censor, R. Davidi, G.T. Herman, Perturbation resilience and superiorization of iterative algorithms, *Inverse Probl.* 26 (2010), 065008.
- [10] Y. Censor, R. Davidi, G.T. Herman, R.W. Schulte, L. Tetrushvili, Projected subgradient minimization versus superiorization, *J. Optim. Theory Appl.* 160 (2014), 730–747.
- [11] Y. Censor, T. Elfving, G. T. Herman, Averaging strings of sequential iterations for convex feasibility problems, In: D. Butnariu, Y. Censor, S. Reich (eds.), *Studies in Computational Mathematics*, 8, pp.101–113, Elsevier, North Holland, Amsterdam, 2001.
- [12] Y. Censor, G.T. Herman, M. Jiang, Guest Editors, Superiorization: Theory and Applications, Special Issue of the journal *Inverse Problems*, Volume 33, Number 4, April 2017.
- [13] Y. Censor, A. J. Zaslavski, Strict Fejér monotonicity by superiorization of feasibility-seeking projection methods, *J. Optim. Theory Appl.* 165 (2015), 172–187.
- [14] Y. Censor, S. A. Zenios, *Parallel Optimization: Theory, Algorithms, and Applications*, Oxford University Press, New York, New York, 1997.
- [15] P. L. Combettes, Hilbertian convex feasibility problem: convergence of projection methods, *Appl. Math. Optim.* 35 (1997), 311–330.
- [16] P. L. Combettes, On the numerical robustness of the parallel projection method in signal synthesis, *IEEE Signal Processing Lett.* 8 (2001), 45–47.
- [17] R. Correa, C. Lemaréchal, Convergence of some algorithms for convex minimization, *Math. Program.* 62 (1993), 261–275.
- [18] V. F. Demyanov, F. Giannessi, V. Karelín, *Optimal control problems and penalization, Nonlinear optimization and related topics Appl. Optim. vol.36*, Kluwer, Dordrecht, 2000.
- [19] V. F. Dem’yanov, L. V. Vasil’ev, *Nondifferentiable Optimization*, Optimization Software, New York, 1985.
- [20] Q.-L. Dong, A. Gibali, D. Jiang, S.-H. Ke, Convergence of projection and contraction algorithms with outer perturbations and their applications to sparse signals recovery, *J. Fixed Point Theory Appl.* 16 (2018), 16.
- [21] F. Giannessi, P. M. Illustrator-Pardalos, T. Rapcsk, *Optimization Theory: Recent Developments from Mátraháza*, Kluwer, 2011.
- [22] F. Giannessi, *Constrained optimization and image space analysis Mathematical Concepts and Methods in Science and Engineering*, 49, Springer, New York, 2005
- [23] A. Gibali, K.-H. Küfer, D. Reem, P. Süß, A generalized projection-based scheme for solving convex constrained optimization problems, *Comput. Optim. Appl.* 70 (2018), 737–762.
- [24] Y. Nesterov, Gradient methods for minimizing composite functions, *Math. Program.* 140 (2013), 125–161.
- [25] Y. Nesterov, Smooth minimization of non-smooth functions, *Math. Program.* 103 (2005), 127–152.
- [26] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course*, Applied Optimization, 87, Kluwer, Boston, Massachusetts, 2004.
- [27] Y. Nesterov, A method for solving the convex programming problem with convergence rate $o(1/k^2)$, *Doklady Akademii Nauk SSSR* 269 (1983), 543–547.
- [28] A. Niemierko, Reporting and analyzing dose distributions: a concept of equivalent uniform dose, *Med. Phys.* 24 (1997), 103–110.
- [29] A. Nedić and D. Bertsekas: Incremental subgradient methods for nondifferentiable optimization, *SIAM J. Optim.* 12 (2001), 109–138.
- [30] A. De Pierro, A. Iusem, A finitely convergent “row-action” method for the convex feasibility problem, *Appl. Math. Optim.* 17 (1988), 225–235.
- [31] G. Di Pillo, F. Giannessi, *Nonlinear Optimization and Related Topics*, Applied Optimization, book series, 36, 2000.
- [32] B. T. Polyak, A general method of solving extremum problems, *Soviet Math. Doklady* 8 (1967), 593–597.
- [33] H. Romeijn, J. Dempsey, J. Li, A unifying framework for multicriteria fluence map optimization models, *Phys. Med. Biol.* 49 (2004), 1991–2013.
- [34] S. Salzo, S. Villa, Inexact and accelerated proximal point algorithms, *J. Convex Anal.* 19 (2012), 1167–1192.
- [35] D. M. Shepard, M. C. Ferris, G. H. Olivera, T. R. Mackie, Optimizing the delivery of radiation therapy to cancer patients, *SIAM Rev.* 419 (1999), 721–744.

- [36] P. Tseng: On accelerated proximal gradient methods for convex-concave optimization, <https://www.mit.edu/~dimitrib/PTseng/papers/apgm.pdf> (May 21, 2008)
- [37] X. Wang, Method of steepest descent and its applications, *IEEE Microwave Wireless Components Letters* 12 (2008), 24-26.
- [38] X. Zhao, K. F. Ng, C. Li, J.-C. Yao, Linear regularity and linear convergence of projection-based methods for solving convex feasibility problems, *Appl. Math. Optim.* 78 (2018), 613-641.