Barzilai-Borwein method with variable sample size for stochastic linear complementarity problems *

Nataša Krejić† Nataša Krklec Jerinkić † Sanja Rapajić †

February 25, 2015

Abstract

A smoothing method for solving stochastic linear complementarity problems (SLCP) is proposed. The expected residual minimization reformulation of the problem is considered and it is approximated by the Sample Average Approximation (SAA). The proposed method is based on sequential solving of a sequence of smoothing problems where each of the smoothing problems is defined with its own sample average approximation. A nonmonotone line search with a variant of the Barzilai-Borwein (BB) gradient direction is used for solving each of the smoothing problems. The BB search direction is efficient and low cost, particularly suitable for nonmonotone line search procedure. The variable sample size scheme allows the sample size to vary across the iterations and the method tends to use smaller sample size far away from the solution. The key point of this strategy is a good balance between the variable sample size strategy, the smoothing sequence and nonmonotonicity. Eventually, the maximal sample size is used and the SAA problem is solved. Presented numerical results indicate that the proposed strategy reduces the overall computational cost.

Key words. Stochastic linear complementarity problems, sample average approximation, variable sample size, nonsmooth systems, nonmonotone line search
AMS subject classifications. 90C33, 65H10

1 Introduction

The stochastic linear complementarity problem (SLCP) consists of finding a vector $x \in R^n$ such that

$$x \geq 0,\ M(\omega)x + q(\omega) \geq 0,\ x^\top (M(\omega)x + q(\omega)) = 0, \quad \omega \in \Omega,$$

*Research supported by Serbian Ministry of Education, Science and Technological Development, grant no. 174030

†Department of Mathematics and Informatics, Faculty of Sciences, University of Novi Sad, Trg Dositeja Obradovića 4, 21000 Novi Sad, Serbia, e-mail: natasak@uns.ac.rs, natasa.krklec@dmi.uns.ac.rs,sanja@dmi.uns.ac.rs.

where Ω is underlying sample space and $M(\omega) \in R^{n,n}$ and $q(\omega) \in R^n$ for each ω.

One way of dealing with SLCP, presented in Chen, Fukushima [3], is considering its Expected Residual Minimization (ERM) reformulation of the form

$$f(x) = E(\|\Phi_\omega(x)\|^2) \to \min, \quad x \geq 0, \tag{1}$$

where

$$\Phi(x,\omega) = (\phi_1, \ldots, \phi_n)^T, \ \phi_i = \phi(x_i, [M(\omega)x]_i + q_i(\omega)), \ i = 1, ..., n, \tag{2}$$

$\Phi(x,\omega) : R^n \times \Omega \to R^n$ and $\phi : R^2 \to R$ is an NCP function. In this paper, we focus on the ERM reformulation based on the min function $\phi(a,b) = \min\{a,b\}$, where

$$\Phi_\omega(x) := \Phi(x,\omega) = \min\{x, M(\omega)x + q(\omega)\}.$$

This is nonconvex, nonsmooth constrained optimization problem. It is shown in Lemma 2.2 Chen et al. [5] that this problem always has a solution if Ω is a finite set.

The function $f(x)$ in ERM reformulation is in the form of mathematical expectation and in general it is rather difficult to compute it accurately. Sample Average Approximation (SAA) is usually employed for estimating $f(x)$. Assume that $\{\omega^1, \omega^2, ..., \omega^{N_{\max}}\}$ from Ω is a sample of random vectors that are independent and identically distributed. Then function $f(x)$ from (1) can be approximated by Monte Carlo sampling with

$$\hat{f}_{N_{\max}}(x) = \frac{1}{N_{\max}} \sum_{j=1}^{N_{\max}} \|\Phi_{\omega^j}(x)\|^2. \tag{3}$$

Thus we consider the following nonsmooth problem

$$\hat{f}_{N_{\max}}(x) \to \min, \quad x \geq 0. \tag{4}$$

One of the common ways for solving nonsmooth problems like (4) is to apply the smoothing technique which consists of considering a sequence of smoothing functions

$$\hat{f}_{N_{\max}}(x, \mu_k), \quad \mu_k > 0,$$

instead of the objective function $\hat{f}_{N_{\max}}(x)$. More precisely, at each iteration, the objective function $\hat{f}_{N_{\max}}(x)$ is approximated by a smooth function $\hat{f}_{N_{\max}}(x, \mu_k)$ with a fixed smoothing parameter $\mu_k > 0$. The sequence of smoothing functions $\hat{f}_{N_{\max}}(x, \mu_k)$ tends to the nonsmooth objective function $\hat{f}_{N_{\max}}(x)$ when $\mu_k \to 0$.

If $N_{\max}$ is a large integer the SAA approximation (3) is expensive and considering the smoothing function $\hat{f}_{N_{\max}}(x, \mu_k)$ in each iteration might be too expensive. One possible remedy is to use a variable sample size strategy i.e. to consider an approximate objective function $\hat{f}_{N_k}(x, \mu_k)$ at the k-th iteration with $N_k \leq N_{\max}$. Balancing the progress in function decrease with the precision of the objective function approximation i.e. using smaller sample when we are far

away from the solution and larger sample when close to the solution, one can reduce the computational cost of solving (4). The key point of variable sample size strategies is to ensure that $N_{\max}$ sample size is used at the final stage of the optimization procedure so that (4) is eventually solved but with smaller cost. The objective function $\hat{f}_{N_k}$ depends on the smoothing parameter μ_k as well, so one needs to define a proper update for the smoothing parameters as well. Clearly, the smoothing sequence should not converge to zero too fast as there is no need to work with high precision if we are far away form the solution and if N_k is significantly smaller than N_{max}. On the other hand, when the final sample size is reached, iterations become costly and one should be able to reduce the smoothing parameters according to the progress made in approaching the stationary point. Therefore keeping these two values, the sample size and the smoothing parameter in a good balance is the key moment in reducing the cost of solving (4).

In this paper we design a variable sample size strategy for the above explained sequence of smoothing problems. The sample size change is based on ideas from Krejić, Krklec [11], [12]. This strategy allows us to increase and decrease the sample size during the optimization process but the presence of smoothing parameters is an additional challenge that is resolved in this paper. The sample size strategy is embedded with an update of the smoothing parameters in such a way that the smoothing parameters allow an increase of precision in the objective function approximation if needed, and follow the progress in approaching the stationary point if the maximal sample size is reached. So, the proposed method at each iteration has its smoothing parameter μ_k and the sample size N_k while the objective function $\hat{f}_{N_{\max}}(x)$ is approximated by $\hat{f}_{N_k}(x, \mu_k)$.

Besides the smoothing parameter and the sample size, at each iteration one needs to specify a search direction. We consider here a modfication of the Barzilai-Borwein (BB) gradient direction [1], which is a low cost and efficient in nonmonotone line search procedures [8], [14].

It is proved that after a finite number of iterations the maximal sample size $N_{\max}$ is reached and kept until the end so the SAA function $\hat{f}_{N_{\max}}(x, \mu_k)$ is minimized. After that we show that the sequence of smoothing parameters tends to zero, which implies that the sequence of smoothing functions tends to the nonsmooth objective function $\hat{f}_{N_{\max}}(x)$. Finally, we prove that any accumulation point of the sequence generated by the method is a Clarke stationary point of $\hat{f}_{N_{\max}}(x)$. Thus, the classical results of smoothing methods for SAA problems are obtained but with significantly smaller cost.

This paper is organized as follows. Some basic definitions and smoothing properties are given in Section 2. The algorithm is proposed in Section 3 and convergence results are presented in Section 4. Numerical experiments are given in the last section, comparing the method proposed in this paper with the method from Li et al. [9].

2 Preliminaries

Throughout the paper $\|\cdot\|$ represents the Euclidian norm, $R^n_+ = \{x \in R^n, x \geq 0\}$, $R^n_{++} = \{x \in R^n, x > 0\}$ and it is assumed that $M(\omega)$ and $q(\omega)$ are measurable functions of ω such that

$$E(\|M(\omega)\|^2 + \|q(\omega)\|^2) < \infty.$$

For continuously differentiable mapping $H : R^n \to R^n$ the Jacobian of H at x is denoted by $H'(x)$, whereas for smooth mapping $g : R^n \to R$ we denote by $\nabla g(x)$ the gradient of g at x and the i-th component of gradient vector $\nabla g(x)$ is denoted by $[\nabla g(x)]_i$. For a given matrix $A \in R^{n,n}$ and a nonempty set of matrices $\mathcal{A} \in R^{n,n}$, the distance between A and $\mathcal{A}$ is $dist(A, \mathcal{A}) = \inf_{B \in \mathcal{A}} \|A - B\|$, the i-th row of matrix A is denoted by $[A]_i$ and e^i, $i = 1, ..., n$ is the canonical base of R^n .

For locally Lipschitzian mapping $H : R^n \to R^n$, the generalized Jacobian of H at x, defined by Clarke [6], is denoted by $\partial H(x)$. Let $\partial_C H(x)$ be the C-generalized Jacobian of H at x defined by

$$\partial_C H(x) = \partial[H(x)]_1 \times \partial[H(x)]_2 \times \ldots \times \partial[H(x)]_n.$$

For locally Lipschitzian mapping $g : R^n \to R$, according to Theorem 2.5.1 of Clarke [6], the generalized gradient of g at x is defined by

$$\partial g(x) = conv\{\lim_{x^k \to x} \nabla g(x^k), x^k \in D_g\},$$

where *conv* represents convex hull and D_g is the subset of R^n where g is differentiable.

Definition 1. *[15] Function $\tilde{g} : R^n \times R_+ \to R$ is a smoothing function of locally Lipschitzian function g if $\tilde{g}(\cdot, \mu)$ is continuously differentiable in R^n for any $\mu \in R_{++}$ and for any $x \in R^n$,*

$$\lim_{z \to x, \mu \to 0} \tilde{g}(z, \mu) = g(x)$$

and $\{\lim_{z \to x, \mu \to 0} \nabla \tilde{g}(z, \mu)\}$ is nonempty and bounded.

The corresponding smoothing problem for the ERM reformulation has been introduced in Li et al. [9] and Zhang, Chen [15] and is defined for a smoothing parameter $\mu > 0$.

A smoothing approximation for the min function defined in Chen, Mangasarian [2] is

$$\phi(a, b, \mu) = \begin{cases} b, & \text{if } \ a - b \geq \frac{\mu}{2} \\ a - \frac{1}{2\mu}(a - b + \frac{\mu}{2})^2, & \text{if } \ -\frac{\mu}{2} < a - b < \frac{\mu}{2} \\ a, & \text{if } \ a - b \leq -\frac{\mu}{2}. \end{cases} \tag{5}$$

A smoothing function for $\Phi_\omega(x)$ defined by (2) is $\tilde{\Phi}_\omega(x, \mu)$, whose components

$$\phi\left(x_i, [M(\omega)x]_i + q_i(\omega), \mu\right), \ \ i = 1, ..., n \tag{6}$$

are defined with (5). It is proved in Zhang, Chen [15] that

$$\tilde{f}(x,\mu) = E(\|\tilde{\Phi}_\omega(x,\mu)\|^2) \tag{7}$$

is a smoothing function for the objective function $f(x)$ in (1).

Using the SAA method, the objective function $f(x)$ from (1) is estimated with $\hat{f}_{N_{\max}}(x)$ from (3) so $\tilde{f}(x,\mu)$ can be approximated by

$$\hat{f}_{N_{\max}}(x,\mu) = \frac{1}{N_{\max}} \sum_{j=1}^{N_{\max}} \|\tilde{\Phi}_{\omega^j}(x,\mu)\|^2, \tag{8}$$

which is a smoothing function for $\hat{f}_{N_{\max}}(x)$.

The algorithm proposed in this paper uses line search with variable sample size, so at the k-th iteration it considers the differentiable function

$$\hat{f}_{N_k}(x,\mu_k) = \frac{1}{N_k} \sum_{j=1}^{N_k} \|\tilde{\Phi}_{\omega^j}(x,\mu_k)\|^2, \tag{9}$$

where $N_k \leq N_{max}$ is the sample size and μ_k is the smoothing parameter.

Since $\|\tilde{\Phi}_{\omega^j}(x,\mu)\|^2 \in \mathcal{C}^1(R^n)$ and $\|\tilde{\Phi}_{\omega^j}(x,\mu)\|^2 \geq 0$, for every $\omega^j \in \Omega$, $\mu \in R_{++}$, $x \in R^n$, it is easy to see that $\hat{f}_N(x,\mu) \in \mathcal{C}^1(R^n)$ and

$$\hat{f}_N(x,\mu) \geq 0, \tag{10}$$

for every $x \in R^n$, $N \in \{1,2,...,N_{\max}\}$.

Let $f(x)$, $\Phi_\omega(x)$, $\hat{f}_{N_{\max}}(x)$ be the functions defined by (1), (2), (3) and $\tilde{\Phi}_\omega(x,\mu)$, $\tilde{f}(x,\mu)$, $\hat{f}_{N_{\max}}(x,\mu)$ be defined by (5)-(7) and (8) respectively. First we will give some properties of these functions which are necessary for the convergence analysis presented later on.

Lemma 1. *[15] Let $\partial\Phi_\omega(x)$ be the generalized Jacobian of $\Phi_\omega(x)$ and $\partial f(x)$ be the generalized gradient of $f(x)$. Denote $\tilde{\kappa} = \frac{1}{4}\sqrt{n}$. For any $\omega \in \Omega$ and $\mu \in R_{++}$ there hold*

a) $\|\tilde{\Phi}_\omega(x,\mu) - \Phi_\omega(x)\| \leq \tilde{\kappa}\mu, \quad x \in R^n,$

b) $\lim_{\mu\to 0} \tilde{\Phi}'_\omega(x,\mu) \in \partial\Phi_\omega(x), \quad x \in R^n,$

c) $\lim_{\mu\to 0} \nabla\tilde{f}(x,\mu) \in \partial f(x), \quad x \in R^n_+,$

d) $\|\tilde{\Phi}'_\omega(x,\mu)\| \leq 2 + \|M(\omega)\|, \quad x \in R^n.$

Since $\partial\Phi_\omega(x) \subseteq \partial_C\Phi_\omega(x)$, Lemma 1 b) implies that $\tilde{\Phi}_\omega(x,\mu)$ has the Jacobian consistency property defined in Chen et al. [4] which means

$$\lim_{\mu\to 0} \tilde{\Phi}'_\omega(x,\mu) \in \partial_C\Phi_\omega(x). \tag{11}$$

The following lemma is an immediate consequence of (11) and the definitions of $\hat{f}_{N_{\max}}(x)$ and $\hat{f}_{N_{\max}}(x,\mu)$ so the proof is omitted.

Lemma 2. *Let $x \in R^n$. Then* $\lim_{\mu \to 0} \nabla \hat{f}_{N_{\max}}(x, \mu) \in \partial \hat{f}_{N_{\max}}(x)$.

The above lemma implies that for every fixed $\delta_1 > 0$, there exists a threshold value $\bar{\mu}(x, \delta_1) > 0$ such that

$$dist(\nabla \hat{f}_{N_{\max}}(x, \mu), \partial \hat{f}_{N_{\max}}(x)) \leq \delta_1, \tag{12}$$

for all $0 < \mu \leq \bar{\mu}(x, \delta_1)$. The exact value of this threshold is necessary for the smoothing parameter update in the algorithm we are proposing and it can be determined as follows.

Since (11) holds by Lemma 1, there follows that for every fixed $\delta > 0$ there exists a threshold value $\bar{\mu}(x, \delta) > 0$ such that

$$dist(\tilde{\Phi}'_\omega(x, \mu), \partial_C \Phi_\omega(x)) \leq \delta, \tag{13}$$

for every $0 < \mu \leq \bar{\mu}(x, \delta)$. The precise definition of $\bar{\mu}(x, \delta)$ is given in the following lemma.

Lemma 3. *Let $x \in R^n$ and $\omega \in \Omega$ be arbitrary but fixed and assume that x is not a solution of the SLCP. Define $\gamma := \max_i \|[M(\omega)]_i - e^i\|$ and*

$$\xi(x) := \min_{i \notin \beta(x)} |x_i - [M(\omega)x]_i - q_i(\omega)|,$$

where $\beta(x) := \{i, \quad x_i = [M(\omega)x]_i + q_i(\omega)\}$. Let $\delta > 0$ be given and define the threshold value

$$\bar{\mu}(x, \delta) := \begin{cases} \frac{2\sqrt{n}\gamma\xi(x)}{\sqrt{n}\gamma - 2\delta}, & \gamma \neq 0 \ \text{ and } \ (\frac{1}{2} - \frac{\delta}{\sqrt{n}\gamma}) > 0 \\ 1, & \text{otherwise.} \end{cases}$$

Then for all $\mu \in (0, \bar{\mu}(x, \delta)]$ the following holds

$$\inf_{V \in \partial_C \Phi_\omega(x)} \sum_{i=1}^{n} \|[\tilde{\Phi}'_\omega(x, \mu)]_i - [V]_i\|^2 \leq \delta^2. \tag{14}$$

Proof. It is sufficient to show that for every $i \in \{1, ..., n\}$ there holds

$$\|[\tilde{\Phi}'_\omega(x, \mu)]_i - [V]_i\| \leq \frac{\delta}{\sqrt{n}}, \tag{15}$$

where $V \in \partial_C \Phi_\omega(x)$, i.e. $[V]_i \in \partial[\Phi_\omega(x)]_i$. For $r_i = [M(\omega)x]_i + q_i(\omega)$ and $y_i = \frac{1}{\mu}(x_i - r_i + \mu/2)$ we obtain

$$[V]_i = \begin{cases} [M(\omega)]_i, & \text{if} \quad x_i > r_i \\ \lambda[M(\omega)]_i + (1 - \lambda)e^i, & \text{if} \quad x_i = r_i, \ \lambda \in [0, 1] \\ e^i, & \text{if} \quad x_i < r_i, \end{cases}$$

$$[\tilde{\Phi}'_\omega(x, \mu)]_i = \begin{cases} [M(\omega)]_i, & \text{if} \quad x_i - r_i \geq \mu/2 \\ y_i[M(\omega)]_i + (1 - y_i)e^i, & \text{if} \quad |x_i - r_i| < \mu/2 \\ e^i & \text{if} \quad x_i - r_i \leq -\mu/2. \end{cases}$$

Notice that $|x_i - r_i| \geq \mu/2$ implies $[\tilde{\Phi}'_\omega(x,\mu)]_i = [V]_i$ and therefore (15) holds trivially. On the other hand, suppose that $|x_i - r_i| < \mu/2$. If $x_i = r_i$ then $y_i = 1/2$ and therefore $\|[\tilde{\Phi}'_\omega(x,\mu)]_i - [V]_i\| = |\lambda - 1/2|\|[M(\omega)]_i - e^i\| = 0$ for $\lambda = 1/2$. This leaves us with the following two remaining cases.

Case 1. If $|x_i - r_i| < \mu/2$ and $x_i < r_i$ we have

$$\|[\tilde{\Phi}'_\omega(x,\mu)]_i - [V]_i\| = |y_i|\|[M(\omega)]_i - e^i\| \leq \frac{1}{\mu}(x_i - [M(\omega)x]_i - q_i(\omega) + \mu/2)\gamma.$$

Therefore, we want to show that for all $0 < \mu \leq \bar{\mu}(x,\delta)$

$$\frac{1}{\mu}(x_i - [M(\omega)x]_i - q_i(\omega) + \mu/2)\gamma \leq \frac{\delta}{\sqrt{n}}. \tag{16}$$

Case 2. If $|x_i - r_i| < \mu/2$ and $x_i > r_i$ we obtain

$$\|[\tilde{\Phi}'_\omega(x,\mu)]_i - [V]_i\| = |y_i - 1|\|[M(\omega)]_i - e^i\| \leq (1 - \frac{1}{\mu}(x_i - [M(\omega)x]_i - q_i(\omega) + \mu/2))\gamma$$

and we want to show that for all $0 < \mu \leq \bar{\mu}(x,\delta)$

$$(1 - \frac{1}{\mu}(x_i - [M(\omega)x]_i - q_i(\omega) + \mu/2))\gamma \leq \frac{\delta}{\sqrt{n}}. \tag{17}$$

If $\gamma = 0$ then (16) and (17) hold trivially for every $\mu > 0$. Hence, suppose that $\gamma \neq 0$. If $\frac{1}{2} \leq \frac{\delta}{\sqrt{n}\gamma}$ then (16) and (17) hold for $\mu \leq \bar{\mu}(x,\delta) = 1$. Finally, if $\frac{1}{2} > \frac{\delta}{\sqrt{n}\gamma}$ then the inequalities (16) and (17) hold for every $0 < \mu \leq \frac{\xi(x)2\sqrt{n}\gamma}{\sqrt{n}\gamma - 2\delta} := \bar{\mu}(x,\delta)$, which completes the proof. □

Since $\|A\|^2 \leq \sum_{i=1}^n \|[A]_i\|^2$ for an arbitrary matrix $A \in R^{n,n}$, from the previous lemma there follows that (13) holds for $0 < \mu \leq \bar{\mu}(x,\delta)$.

Let $\delta_1 > 0$. For $\delta(\omega^j) < \frac{\delta_1}{2\|\Phi_{\omega^j}(x)\|}$, $j = 1, ..., N_{\max}$, by Lemma 3 we can obtain values $\bar{\mu}(x,\delta(\omega^j))$, $j = 1, ..., N_{\max}$ and the threshold value

$$\bar{\bar{\mu}}(x,\delta_1) = \min_{j=1,...,N_{\max}} \bar{\bar{\mu}}_j(x,\delta_1),$$

where

$$\bar{\bar{\mu}}_j(x,\delta_1) = \min\{\bar{\mu}(x,\delta(\omega^j)), \frac{4(\frac{\delta_1}{2} - \|\Phi_{\omega^j}(x)\|\delta(\omega^j))}{\sqrt{n}(2 + \|M(\omega^j)\|)}\}.$$

Obviously, for a given $\delta_1 > 0$ and the threshold value $\bar{\bar{\mu}}(x,\delta_1)$ we have that (12) holds for all $0 < \mu \leq \bar{\bar{\mu}}(x,\delta_1)$.

3 The algorithm

In this section we present a variable sample size smoothing algorithm for solving problem (4). The algorithm is based on Barzilai-Borwein gradient method and uses a nonmonotone line search which allows that the sample size varies across the iterations and thus makes the process significantly cheaper.

The sample size scheduling is based on two measures, the lack of precision $\epsilon_\delta^k(x)$ and the decrease in function value, dm_k. These two measures are calculated at each iteration. The lack of precision is due to the difference between the objective function $\hat{f}_{N_{\max}}(x)$ and $\hat{f}_{N_k}(x)$ used in the k-th iteration, while the decrease in function value dm_k approximates $\hat{f}_{N_k}(x^{k+1}) - \hat{f}_{N_k}(x^k)$.

So, at the k-th iteration we are dealing with the sample size N_k, the smoothing parameter μ_k and the smooth function $\hat{f}_{N_k}(x, \mu_k)$. Since $\hat{f}_{N_k}(x, \mu_k)$ is continuously differentiable, x^k is its local minimizer if and only if $\| \min\{x^k, \nabla \hat{f}_{N_k}(x^k, \mu_k)\}\| = 0$, [15].

The smoothing procedure is governed by a sequence of smoothing parameters, which is updated following the principle used for solving nonlinear complementarity problems, [4], [10], [13] but in the framework of the variable sample scheme. Thus the smoothing parameters update changes depending on N_k and $N_{\max}$.

The search direction which is used is a kind of the Barzilai-Borwein (BB) direction used in Li et al. [9] as well. It is defined in the following way

$$d_i^k = \begin{cases} -\frac{1}{\alpha_k}[\nabla \hat{f}_{N_k}(x^k, \mu_k)]_i, & \text{if} \quad i \in I_1(x^k) \\ -\frac{[\nabla \hat{f}_{N_k}(x^k, \mu_k)]_i}{\alpha_k + \frac{[\nabla \hat{f}_{N_k}(x^k, \mu_k)]_i}{x_i^k}}, & \text{if} \quad i \in I_2(x^k) \\ -x_i^k, & \text{if} \quad i \in I_3(x^k) \end{cases}, \tag{18}$$

where $x^k \geq 0$, $\tau > 0$, I_1, I_2 and I_3 are sets of indexes

$$\begin{aligned} I_1(x^k) &= \{i, \ i \in \{1, ..., n\}, \quad [\nabla \hat{f}_{N_k}(x^k, \mu_k)]_i \leq 0\}, \\ I_2(x^k) &= \{i, \ i \in \{1, ..., n\}, \quad [\nabla \hat{f}_{N_k}(x^k, \mu_k)]_i > 0 \quad \text{and} \quad x_i^k > \tau\}, \\ I_3(x^k) &= \{i, \ i \in \{1, ..., n\}, \quad [\nabla \hat{f}_{N_k}(x^k, \mu_k)]_i > 0 \quad \text{and} \quad 0 \leq x_i^k \leq \tau\} \end{aligned}$$

and

$$\alpha_k = \begin{cases} \max\{\alpha_{\min}, \frac{(s^{k-1})^T y^{k-1}}{\|s^{k-1}\|^2}\}, \text{ if } \|s^{k-1}\| > 0 \wedge mod(k, 4) = 0, 1 \\ \max\{\alpha_{\min}, \frac{\|y^{k-1}\|^2}{(s^{k-1})^T y^{k-1}}\}, \text{ if } (s^{k-1})^T y^{k-1} \neq 0 \wedge mod(k, 4) = 2, 3 \\ \alpha_{\min}, \quad \text{else} \end{cases} \tag{19}$$

$\alpha_{\min} > 0$, $\quad s^{k-1} = x^k - x^{k-1}$ and

$$y^{k-1} = \nabla \hat{f}_{N_k}(x^k, \mu_k) - \nabla \hat{f}_{N_{k-1}}(x^{k-1}, \mu_{k-1}). \tag{20}$$

It is easy to see that d^k is feasible and descent for $\hat{f}_{N_k}(x^k, \mu_k)$, because $x^k + d^k \geq 0$ and $\nabla \hat{f}_{N_k}(x^k, \mu_k)^T d^k \leq 0$. Therefore, the monotone line search can also be applied, but we use a nonmonotone search because it allows larger step sizes and combines well with the BB direction.

We can now state the main algorithm as follows. Let $\omega^1, \ldots, \omega^{N_{\max}}$ be given.

ALGORITHM 1.

S0 Input parameters: $N_{\max}, N_0^{\min} \in N,\ x^0 \in R_+^n,\ \bar{\kappa} > 0,\ 0 < \alpha_{\min} \leq \alpha_0 < 1,$ $\eta, \beta \in (0,1)$. Let $\{\varepsilon_k\}_{k\in N}$ be a sequence such that $\varepsilon_k > 0$, $\sum_{k\in N} \varepsilon_k \leq \varepsilon < \infty$.

S1 Set $k = 0$, $N_k = N_0^{\min}$, $\tilde{\beta} = \hat{f}_{N_{\max}}(x^0)$, $\mu_0 = \frac{\alpha\tilde{\beta}}{2\bar{\kappa}}$, $\mu_k = \mu_0$ and $x^k = x^0$.

S2 Compute $\hat{f}_{N_k}(x^k, \mu_k)$, $\nabla\hat{f}_{N_k}(x^k, \mu_k)$ and $\epsilon_\delta^{N_k}(x^k, \mu_k)$ by (21).

S3 If $\| \min\{x^k, \nabla\hat{f}_{N_k}(x^k, \mu_k)\}\| < \mu_k$

1) if $N_k = N_{\max}$ or ($N_k < N_{\max}$ and $\epsilon_\delta^{N_k}(x^k, \mu_k) > 0$) set $N_{k+1} = N_{\max}$ and $N_{k+1}^{\min} = N_{\max}$.

2) if $N_k < N_{\max}$ and $\epsilon_\delta^{N_k}(x^k, \mu_k) = 0$ set $N_{k+1} = N_k + 1$ and $N_{k+1}^{\min} = N_k^{\min} + 1$.

Set $x^{k+1} = x^k$, $\mu_{k+1} = \frac{\mu_k}{2}$, $\alpha_{k+1} = \alpha_{\min}$ and go to step S11.
If $\| \min\{x^k, \nabla\hat{f}_{N_k}(x^k, \mu_k)\}\| \geq \mu_k$, go to step S4.

S4 Determine the BB direction d^k by (18).

S5 Find the smallest nonnegative integer j such that $\nu_k = \beta^j$ satisfies

$$\hat{f}_{N_k}(x^k + \nu_k d^k, \mu_k) \leq \hat{f}_{N_k}(x^k, \mu_k) + \eta\nu_k (d^k)^T \nabla\hat{f}_{N_k}(x^k, \mu_k) + \varepsilon_k.$$

S6 Set $x^{k+1} = x^k + \nu_k d^k$ and $dm_k = -\nu_k (d^k)^T \nabla\hat{f}_{N_k}(x^k, \mu_k)$.

S7 Determine N_k^+ using Algorithm 2 and then determine N_{k+1}.

S8 Determine μ_{k+1} using Algorithm 3.

S9 Determine $N_{k+1}^{\min}$.

S10 Compute α_{k+1} by (19).

S11 Set $k = k + 1$ and go to step S2.

Updating the sample size relays conceptually on algorithms stated in [11] and [12], but the smoothing procedure makes it more complicated as we have here two sources of imprecision - an incomplete sample and a smoothing parameter. Thus the main difference in the algorithms presented here is the need to include the influence of the smoothing parameters. The candidate sample size N_k^+ is determined by comparing the measure of decrease in the objective function dm_k and the so called lack of precision defined by

$$\epsilon_\delta^{N_k}(x^k, \mu_k) = \hat{\sigma}_{N_k}(x_k, \mu_k) \frac{\alpha_\delta}{\sqrt{N_k}}, \tag{21}$$

where

$$\hat{\sigma}^2_{N_k}(x^k,\mu_k)=\frac{1}{N_k-1}\sum_{i=1}^{N_k}(\|\tilde{\Phi}_{\omega^i}(x^k,\mu_k)\|^2-\hat{f}_{N_k}(x^k,\mu_k))^2$$

and α_δ is a quantile of the standard normal distribution. The lack of precision represents the approximate measure of the error bound for $|\hat{f}_{N_k}(x^k,\mu_k)-f(x^k,\mu_k)|$. The main idea is to find the sample size N_k^+ such that

$$dm_k \approx d\ \epsilon_\delta^{N_k^+}(x^k,\mu_k),$$

where $d\in(0,1]$ and $N_k^{\min}\le N_k^+\le N_{\max}$. For example, if the decrease measure is greater than some portion of the lack of precision we are probably far away from the solution. In that case, we do not want to impose high precision and therefore the sample size is decreased if possible. We state the algorithm for choosing the candidate sample size N_k^+.

ALGORITHM 2.

S0 Input parameters: $\tilde{\nu}_1, d\in(0,1),\ dm_k,\ N_k^{min},\ \epsilon_\delta^{N_k}(x^k,\mu_k)$.

S1 Determine N_k^+

1) $dm_k = d\ \epsilon_\delta^{N_k}(x^k,\mu_k) \quad\rightarrow\quad N_k^+=N_k$.

2) $dm_k > d\ \epsilon_\delta^{N_k}(x^k,\mu_k)$
Starting with $N=N_k$, while $dm_k > d\ \epsilon_\delta^{N}(x^k,\mu_k)$ and $N>N_k^{min}$, decrease N by 1 and calculate $\epsilon_\delta^{N}(x^k,\mu_k) \quad\rightarrow\quad N_k^+$.

3) $dm_k < d\ \epsilon_\delta^{N_k}(x^k,\mu_k)$

i) $dm_k \ge \tilde{\nu}_1 d\ \epsilon_\delta^{N_k}(x^k,\mu_k)$
Starting with $N=N_k$, while $dm_k < d\ \epsilon_\delta^{N}(x^k,\mu_k)$ and $N<N_{max}$, increase N by 1 and calculate $\epsilon_\delta^{N}(x^k,\mu_k) \quad\rightarrow\quad N_k^+$.

ii) $dm_k < \tilde{\nu}_1 d\ \epsilon_\delta^{N_k}(x^k,\mu_k) \quad\rightarrow\quad N_k^+=N_{max}$.

After finding the candidate sample size, we perform the safeguard check in order to prohibit the decrease of the sample size which seems to be unproductive. More precisely, if $N_k^+<N_k$ we calculate

$$\rho_k=\left|\frac{\hat{f}_{N_k^+}(x^k)-\hat{f}_{N_k^+}(x^{k+1})}{\hat{f}_{N_k}(x^k)-\hat{f}_{N_k}(x^{k+1})}-1\right|.$$

We do not allow the decrease if the previously stated parameter is relatively large. Namely, if $\rho_k\ge\frac{N_k-N_k^+}{N_k}$ we set $N_{k+1}=N_k$. In all the other cases, the decrease is accepted and $N_{k+1}=N_k^+$. Notice that $N_{k+1}\ge N_k^+$ either way.

Updating the lower sample size bound $N_k^{\min}$ is also very important. This bound is increased only if $N_{k+1}>N_k$ and we have not made big enough decrease of the function $\hat{f}_{N_{k+1}}$ since the last time we started to use it, i.e. if

$$\frac{\hat{f}_{N_{k+1}}(x^{h(k)},\mu_{k+1})-\hat{f}_{N_{k+1}}(x^{k+1},\mu_{k+1})}{k+1-h(k)}<\frac{N_{k+1}}{N_{\max}}\epsilon_\delta^{N_{k+1}}(x^{k+1},\mu_{k+1}),$$

where $h(k)$ is the iteration at which we started to use the sample size N_{k+1} for the last time. In that case, we set $N_{k+1}^{\min} = N_{k+1}$ while in all the other cases the lower bound remains unchanged.

The following algorithm presents the way for updating the smoothing parameter μ_k. The threshold value $\bar{\mu}(x^{k+1}, \gamma\tilde{\beta})$ for the smoothing parameter, in Step S2 of Algorithm 3, is defined in the previous section and it controls the distance between the smooth gradient and the generalized gradient.

ALGORITHM 3.

S0 Input parameters: $\bar{\kappa},\ \gamma > 0,\ \alpha,\ \bar{\xi} \in (0,1),\ N_k,\ N_{k+1},\ \mu_k,\ \tilde{\beta}$.

S1 **1)** If $N_{k+1} = N_k = N_{\max}$ go to step S2.

2) If $N_k < N_{k+1}$ put $\mu_{k+1} = \frac{\mu_k}{2}$ and stop.

3) Else $\mu_{k+1} = \mu_k$ and stop.

S2 If

$$\hat{f}_{N_{\max}}(x^{k+1}) \leq \max\{\bar{\xi}\tilde{\beta}, \frac{|\hat{f}_{N_{\max}}(x^{k+1}) - \hat{f}_{N_{\max}}(x^{k+1}, \mu_k)|}{\alpha}\}$$

then $\tilde{\beta} = \hat{f}_{N_{\max}}(x^{k+1})$ and $\mu_{k+1} \leq \min\{\frac{\mu_k}{2}, \frac{\alpha\tilde{\beta}}{2\bar{\kappa}}, \bar{\mu}(x^{k+1}, \gamma\tilde{\beta})\}$,
else $\mu_{k+1} = \mu_k$.

Let us comment here on the mutual dependence of μ_k and N_k. First of all, μ_k clearly influences N_k^+ and thus N_{k+1}, as $\varepsilon_\delta^N(x^k, \mu_k)$ clearly depends on μ_k. On the other hand the update rule for μ_k depends on N_k as one can see in Algorithm 3. If $N_k < N_{k+1}$ we decrease μ_k as the precision of the approximate objective function is increased so the smoothing parameter should be smaller. Otherwise, if $N_{k+1} \leq N_k$, the smoothing parameter does not change as there is no need for precision increase. Finally, when $N_{\max}$ is reached, the update procedure for μ_k is based on the threshold value and the progress made in the objective function decrease. Furthermore, μ_k determines the tolerance at S3 of Algorithm 1 and thus crucially influences the scheduling sequence.

Algorithm 1 is well defined. The line search in S5 terminates after a finite number of trials as the BB direction is feasible and descent. In fact the non-negative term ε_k makes it well defined even for a direction d^k which is not descent. In the case of BB search direction this term provides additional possibilities for the steplength. Also, notice that the smoothing parameter is updated in step S3 by $\mu_{k+1} = \mu_k/2$ or in step S8 according to Algorithm 3.

4 Convergence analysis

The convergence results are obtained in the three main stages. First, we show that after a finite number of iterations the algorithm ends up at the full sample i.e. $N_k = N_{\max}$ for k large enough. Then, we prove that the sequence of smoothing parameters tends to zero and therefore the sequence of approximate functions tends to the objective function $\hat{f}_{N_{max}}(x)$. Finally, we prove that every

accumulation point of the sequence generated by Algorithm 1 is a Clarke stationary point of $\hat{f}_{N_{max}}(x)$. Moreover, we state the conditions which imply the convergence towards the global optimum. The following assumption is imposed on the random variables.

A 1. *Functions $M(\omega)$, $q(\omega)$ are measurable and $E(\|M(\omega)\|^2 + \|q(\omega)\|^2) < \infty$.*

We begin the analysis by stating the following technical lemma.

Lemma 4. *Let $C \subset R^n$ be a compact set. Then for every $x \in C$, $N \in \{1, 2, ..., N_{\max}\}$ and $\mu, \mu_1, \mu_2 \in R_{++}$ there exists $\bar{\kappa} > 0$ such that following inequalities hold*

a) $|\hat{f}_N(x, \mu) - \hat{f}_N(x)| \leq \bar{\kappa}\mu$,

b) $|\hat{f}_N(x, \mu_2) - \hat{f}_N(x, \mu_1)| \leq \bar{\kappa}|\mu_1 - \mu_2|$.

Proof. a) Since $\|\Phi_{\omega^j}(x)\|$ is continuous and C is assumed to be compact there follows that $\|\Phi_{\omega^j}(x)\|$ is bounded on C. More precisely, there exists constant $M_3(\omega^j) < \infty$ such that

$$\|\Phi_{\omega^j}(x)\| \leq M_3(\omega^j), \tag{22}$$

for every $x \in C$, $\omega^j \in \Omega$. Since $N \in \{1, 2, ..., N_{\max}\}$ is fixed, let $M = \max_{j=1}^{N_{\max}} M_3(\omega^j)$ and choose $\bar{\kappa} = \tilde{\kappa}(\tilde{\kappa}\mu + 2M)$. Then by Lemma 1 a) follows

$$\|\tilde{\Phi}_{\omega^j}(x, \mu)\| \leq \|\tilde{\Phi}_{\omega^j}(x, \mu) - \Phi_{\omega^j}(x)\| + \|\Phi_{\omega^j}(x)\| \leq \tilde{\kappa}\mu + M_3(\omega^j) \tag{23}$$

and (22) and (23) imply

$$\begin{aligned} |\|\tilde{\Phi}_{\omega^j}(x, \mu)\|^2 - \|\Phi_{\omega^j}(x)\|^2| &\leq \|\tilde{\Phi}_{\omega^j}(x, \mu) - \Phi_{\omega^j}(x)\|(\|\tilde{\Phi}_{\omega^j}(x, \mu)\| + \|\Phi_{\omega^j}(x)\|) \\ &\leq \tilde{\kappa}\mu(\tilde{\kappa}\mu + 2M_3(\omega^j)), \end{aligned}$$

for every $\omega^j \in \Omega$. Therefore

$$\begin{aligned} |\hat{f}_N(x, \mu) - \hat{f}_N(x)| &= \frac{1}{N}\sum_{j=1}^{N} |\|\tilde{\Phi}_{\omega^j}(x, \mu)\|^2 - \|\Phi_{\omega^j}(x)\|^2| \\ &\leq \max_{j=1}^{N} |\|\tilde{\Phi}_{\omega^j}(x, \mu)\|^2 - \|\Phi_{\omega^j}(x)\|^2| \\ &\leq \tilde{\kappa}(\tilde{\kappa}\mu + 2M)\mu = \bar{\kappa}\mu, \end{aligned}$$

which completes the proof.
b) It can be proved in a similar way as a). □

The following theorem states that after a finite number of iterations, the sample size becomes stationary with $N_k = N_{\max}$. This result implies that a solution obtained by the proposed algorithm is of the same quality as a solution of the full SAA problem. The proof follows the ideas presented in Krejić, Krklec [11], [12] but the presence of the smoothing parameters makes it technically more demanding. It is included in this paper for the sake of completeness.

Theorem 1. *Suppose that A1 holds and that the sequence $\{x^k\}_{k\in\mathbb{N}}$ generated by Algorithm 1 is bounded. Furthermore, suppose that there are positive constants κ and $n_0 \in N$ such that $\epsilon_\delta^{N_k}(x^k, \mu_k) \geq \kappa$ for every $k \geq n_0$. Then, there exists $q \in N$ such that $N_k = N_{\max}$ for every $k \geq q$.*

Proof. First of all, suppose that $\|\min\{x^k, \nabla \hat{f}_{N_k}(x^k, \mu_k)\}\| < \mu_k$ happens infinitely many times. Then, step S3 of Algorithm 1 will eventually yield $N_k^{\min} = N_{\max}$, which furthermore implies existence of an iteration $q \in \mathbb{N}$ such that $N_k = N_{\max}$ for every $k \geq q$. Therefore, let us consider the case $\|\min\{x^k, \nabla \hat{f}_{N_k}(x^k, \mu_k)\}\| > \mu_k$ for every $k \geq n_1$ where n_1 is some finite integer. Without loss of generality, we can assume that $n_1 > n_0$. Then $\|\nabla \hat{f}_{N_k}(x^k, \mu_k)\| > 0$ and d^k is descent direction for every $k \geq n_1$. Now, let us prove that the sample size can not be stacked at a size lower than $N_{\max}$.

Suppose that there exists $\tilde{n} > n_1$ such that $N_k = N^1 < N_{\max}$ for every $k \geq \tilde{n}$. In that case, Algorithm 3 implies that $\mu_{k+1} = \mu_k = \mu$ for every $k \geq \tilde{n}$. Denoting $g_k = \nabla \hat{f}_{N^1}(x^k, \mu)$, we obtain that for every $k \geq \tilde{n}$

$$\hat{f}_{N^1}(x^{k+1}, \mu) \leq \hat{f}_{N^1}(x^k, \mu) + \varepsilon_k + \eta \nu_k (d^k)^T g_k.$$

Furthermore, by using the induction argument, the summability of the sequence $\{\varepsilon_k\}$ and the inequality (10), we obtain that

$$\lim_{k\to\infty} dm_k = \lim_{j\to\infty} -\nu_{\tilde{n}+j} (\nabla \hat{f}_{N^1}(x^{\tilde{n}+j}, \mu))^T d^{\tilde{n}+j} = 0.$$

On the other hand, we have that $\epsilon_\delta^{N^1}(x^k, \mu) \geq \kappa > 0$ for every $k \geq n_0$, which implies that $\tilde{\nu}_1 d\ \epsilon_\delta^{N^1}(x^k, \mu)$ is bounded from below for all k sufficiently large. Therefore, there exists at least one p such that $dm_p < \tilde{\nu}_1 d\ \epsilon_\delta^{N^1}(x^p, \mu)$. This implies $N_{p+1} \geq N_p^+ = N_{\max}$ which is contrary to the current assumption that the sample size stays at N^1. Therefore, the remaining two possible situations are as follows: there exists $\tilde{n}$ such that $N_k = N_{\max}$ for every $k \geq \tilde{n}$ or the sequence of sample sizes oscillates.

Let us suppose that the sequence of sample sizes oscillates. Notice that the existence of $\bar{j} \in \mathbb{N}$ such that $N_{\bar{j}}^{\min} = N_{\max}$ would imply the first of the considered two situations. Therefore, we have that $N_k^{\min} < N_{\max}$ for every $k \in \mathbb{N}$. Furthermore, this implies that the signal for increasing $N_k^{\min}$ could come only finitely many times and we conclude that there exists an iteration $r \geq n_1$ such that for every $k \geq r$ we have one of the following scenarios:

M1 $N_{k+1} \leq N_k$

M2 $N_{k+1} > N_k$ and we have enough decrease in $\hat{f}_{N_{k+1}}$

M3 $N_{k+1} > N_k$ and we did not use the sample size N_{k+1} before.

Now, let $\bar{N}$ be the maximal sample size that is used at infinitely many iterations. Furthermore, define the set of iterations $\bar{K}_0$ at which sample size increases on $\bar{N}$ and set $\bar{K} = \bar{K}_0 \bigcap \{r, r+1, \ldots\}$. Notice that $N_k < N_{k+1} = \bar{N}$

for every $k \in \bar{K}$. This implies that every iteration in $\bar{K}$ excludes the scenario M1. Moreover, without loss of generality, we can say that scenario M3 is the one that can also be excluded. This leads us to the conclusion that M2 is the only possible scenario for iterations in $\bar{K}$. Therefore, for every $k \in \bar{K}$ the following is true

$$\hat{f}_{\bar{N}}(x^{h(k)}, \mu_{k+1}) - \hat{f}_{\bar{N}}(x^{k+1}, \mu_{k+1}) \geq \frac{\bar{N}}{N_{\max}}(k+1-h(k))\epsilon_{\delta}^{\bar{N}}(x^{k+1}, \mu_{k+1}) \geq \frac{\kappa}{N_{\max}}.$$

Define $S := \kappa / N_{\max}$. We know that S is a positive constant. Define also a subsequence of iterations $\{x^{s_j}\}_{j \in \mathbb{N}} := \{x^k\}_{k \in \bar{K}}$. Recall that $h(k)$ defines the iteration at which we started to use the sample size $\bar{N}$ for the last time before the iteration $k+1$. Having all this in mind, we know that for every j

$$\hat{f}_{\bar{N}}(x^{s_{j+1}}, \mu_{s_{j+1}}) \leq \hat{f}_{\bar{N}}(x^{s_j}, \mu_{s_{j+1}}) - S.$$

Furthermore, we know that sequence $\{\mu_k\}_{k \in \mathbb{N}}$ is nonincreasing which together with Lemma 4 implies that for every i, j and x

$$\hat{f}_{\bar{N}}(x, \mu_{i+j}) \leq \hat{f}_{\bar{N}}(x, \mu_i) + \bar{\kappa}(\mu_i - \mu_{i+j}).$$

Again, using the induction argument and the previous two inequalities we obtain that for every $j \in \mathbb{N}$

$$0 \leq \hat{f}_{\bar{N}}(x^{s_j}, \mu_{s_j}) \leq \hat{f}_{\bar{N}}(x^{s_0}, \mu_{s_0}) + \bar{\kappa}(\mu_{s_0} - \mu_{s_j}) - jS \leq \hat{f}_{\bar{N}}(x^{s_0}, \mu_{s_0}) + \bar{\kappa}\mu_{s_0} - jS.$$

Letting $j \to \infty$, we obtain the contradiction and therefore we conclude that the only possible situation is that there exists an iteration $\tilde{n}$ such that $N_k = N_{\max}$ for every $k \geq \tilde{n}$. $\square$

Remark. The previous theorem is valid if the iterative sequence is bounded. This assumption holds whenever the matrix $M(\omega)$ is a stochastic R_0 matrix (see [5]). In that case we have that the level sets of the original objective function are bounded and thus the level sets of $\hat{f}_{N_k}$ are also bounded for each $N_k \in \{N_0, \ldots, N_{\max}\}$. Therefore the sequence $\{x^k\}$ belongs to the union of finitely many bounded sets. For details on level sets one can see [13].

Next, we prove that the sequence of smoothing parameters converges to zero. This results is essentially important since it provides that the sequence of smoothing functions converges to the original objective function $\hat{f}_{N_{max}}(x)$. Let us recall that $q \in \mathbb{N}$ is such that $N_k = N_{\max},\ k \geq q$.

Theorem 2. *Suppose that the assumptions of Theorem 1 are satisfied. Then*

$$\lim_{k \to \infty} \mu_k = 0.$$

Proof. Define S and K in the following way

$$K = \{k \in \mathbb{N} : k > q,\ \hat{f}_{N_{\max}}(x^k) \leq \max\{\bar{\xi}\tilde{\beta}, \frac{|\hat{f}_{N_{\max}}(x^k, \mu_{k-1}) - \hat{f}_{N_{\max}}(x^k)|}{\alpha}\}\}. \tag{24}$$

$$S = \{k \in \mathbb{N} : \ k > q, \ \| \min\{x^k, \nabla \hat{f}_{N_{max}}(x^k, \mu_k)\}\| < \mu_k\}. \tag{25}$$

Notice that for every $k > q$ the smoothing parameter μ_k is decreased only if $k \in S \cup K$. Suppose that there exists $\bar{k} > q$ such that $\mu_k = \mu > 0$ for every $k \geq \bar{k}$. This implies that S is finite as well as K and for every $k \geq \bar{k}$

$$\| \min\{x^k, \nabla \hat{f}_{N_{max}}(x^k, \mu)\}\| \geq \mu > 0. \tag{26}$$

Moreover, the line search becomes

$$\hat{f}_{N_{\max}}(x^{k+1}, \mu) \leq \hat{f}_{N_{\max}}(x^k, \mu) + \eta \nu_k (d^k)^T \nabla \hat{f}_{N_{\max}}(x^k, \mu) + \varepsilon_k.$$

By the induction argument we obtain that for every $j \in \mathbb{N}$

$$0 \leq \hat{f}_{N_{\max}}(x^{\bar{k}+j}, \mu) \leq \hat{f}_{N_{\max}}(x^{\bar{k}}, \mu) + \eta \sum_{i=0}^{j-1} \nu_{\bar{k}+i} (d^{\bar{k}+i})^T \nabla \hat{f}_{N_{\max}}(x^{\bar{k}+i}, \mu) + \sum_{i=0}^{j-1} \varepsilon_{\bar{k}+i}$$

and therefore

$$0 \leq -\sum_{i=0}^{\infty} \nu_{\bar{k}+i} (d^{\bar{k}+i})^T \nabla \hat{f}_{N_{\max}}(x^{\bar{k}+i}, \mu) \leq (\hat{f}_{N_{\max}}(x^{\bar{k}}, \mu) + \varepsilon)/\eta < \infty.$$

Clearly,

$$\lim_{k \to \infty} \nu_k (d^k)^T \nabla \hat{f}_{N_{\max}}(x^k, \mu) = 0. \tag{27}$$

Let us now show that $\lim_{k\to\infty} (d^k)^T \nabla \hat{f}_{N_{\max}}(x^k, \mu) = 0$. Suppose that there exists $S_1 \subseteq \mathbb{N}$ such that

$$|(d^k)^T \nabla \hat{f}_{N_{\max}}(x^k, \mu)| \geq \tilde{\varepsilon} > 0 \tag{28}$$

for every $k \in S_1$. In that case (27) implies that $\lim_{k \in S_1} \nu_k = 0$ and therefore there exists $S_2 \subseteq S_1$ such that for every $k \in S_2$

$$\hat{f}_{N_{\max}}(x^k + \nu'_k d^k, \mu) > \hat{f}_{N_{\max}}(x^k, \mu) + \eta \nu'_k (d^k)^T \nabla \hat{f}_{N_{\max}}(x^k, \mu),$$

where $\nu'_k = \nu_k / \beta$. Now, by the Mean Value Theorem we obtain that for every $k \in S_2$ there exists $t_k \in (0, 1)$ such that

$$(d^k)^T \nabla \hat{f}_{N_{\max}}(x^k + t_k \nu'_k d^k, \mu) > \eta (d^k)^T \nabla \hat{f}_{N_{\max}}(x^k, \mu).$$

Since $\{x^k\}_{k \in \mathbb{N}}$ is bounded by the theorem assumptions, it follows that there is a sequence $S_3 \subseteq S_2$ such that $\lim_{k \in S_3} (x^k, d^k) = (\tilde{x}, \tilde{d})$. Letting $k \to \infty, k \in S_3$ in the previous inequality we obtain $(\tilde{d})^T \nabla \hat{f}_{N_{\max}}(\tilde{x}, \mu) \geq \eta (\tilde{d})^T \nabla \hat{f}_{N_{\max}}(\tilde{x}, \mu)$. Since $(d^k)^T \nabla \hat{f}_{N_k}(x_k, \mu_k) \leq 0$ for every k and $\eta \in (0, 1)$ the previous inequality implies that

$$0 = (\tilde{d})^T \nabla \hat{f}_{N_{\max}}(\tilde{x}, \mu) = \lim_{k \in S_3} (d^k)^T \nabla \hat{f}_{N_{\max}}(x^k, \mu)$$

which is in contradiction with (28). Therefore, we conclude that

$$\lim_{k\to\infty}(d^k)^T\nabla\hat{f}_{N_{\max}}(x^k,\mu)=0. \tag{29}$$

The definition of d^k implies that for every $k>q$

$$\begin{aligned}(d^k)^T\nabla\hat{f}_{N_{\max}}(x^k,\mu) &= -\sum_{j\in I_1(x^k)}\frac{1}{\alpha_k}[\nabla\hat{f}_{N_{\max}}(x^k,\mu)]_i^2-\sum_{j\in I_2(x^k)}\frac{[\nabla\hat{f}_{N_{\max}}(x^k,\mu)]_i^2}{\alpha_k+\frac{[\nabla\hat{f}_{N_{\max}}(x^k,\mu)]_i}{x_i^k}}\\ &\quad -\sum_{j\in I_3(x^k)}x_i^k[\nabla\hat{f}_{N_{\max}}(x^k,\mu)]_i.\end{aligned}$$

Given that α_k is bounded and that each of the three sums above is non-negative, (29) implies

$$\lim_{k\to\infty}[\nabla\hat{f}_{N_{\max}}(x^k,\mu)]_i=0,\quad i\in I_1(x^k)\cup I_2(x^k)$$

and

$$\lim_{k\to\infty}x_i^k[\nabla\hat{f}_{N_{\max}}(x^k,\mu)]_i=0,\quad i\in I_3(x^k).$$

Thus, $\lim_{k\to\infty}\min\{x_i^k,[\nabla\hat{f}_{N_{\max}}(x^k,\mu)]_i\}=0$ for every $i\in\{1,2,\ldots,n\}$ and

$$\lim_{k\to\infty}\|\min\{x^k,\nabla\hat{f}_{N_{\max}}(x^k,\mu)\}\|=0.$$

This is in contradiction with (26) and we conclude that $\lim_{k\to\infty}\mu_k=0$. □

Before proving the main result, we state the following lemma. Recall that $x^*\geq 0$ is a Clarke stationary point of locally Lipschitzian function $\hat{f}_{N_{\max}}(x)$ if there exists $V\in\partial\hat{f}_{N_{\max}}(x^*)$ such that $V^T(x-x^*)\geq 0$ for every $x\geq 0$. Moreover, we know that $\partial\hat{f}_{N_{\max}}(x^*)$ contains the set

$$G_{\hat{f}_{N_{\max}}}(x^*)=\{\lim_{x^k\to x^*,\mu_k\to 0}\nabla\hat{f}_{N_{\max}}(x^k,\mu_k)\}$$

which is nonempty and bounded [15]. Therefore, in order to prove that $x^*\geq 0$ is a Clarke stationary point of function $\hat{f}_{N_{\max}}(x)$, it is sufficient to prove the existence of $V\in G_{\hat{f}_{N_{\max}}}(x^*)$ such that for every $x\geq 0$ the following holds

$$V^T(x-x^*)\geq 0.$$

Lemma 5. *If* $\lim_{k\to\infty}\|\min\{x^k,\nabla\hat{f}_{N_{\max}}(x^k,\mu_k)\}\|=0$ *and* $\lim_{k\to\infty}\mu_k=0$ *then every accumulation point* $x^*\geq 0$ *of the sequence* $\{x^k\}$ *is a Clarke stationary point of* $\hat{f}_{N_{max}}(x)$.

Proof. Let $x^*\geq 0$ be an arbitrary accumulation point of the sequence $\{x^k\}$, i.e. let $L\subseteq\mathbb{N}$ be the sequence such that $x^*=\lim_{k\in L}x^k$. Then $\lim_{k\to\infty}\|\min\{x^k,\nabla\hat{f}_{N_{\max}}(x^k,\mu_k)\}\|=0$ implies

$$0=\lim_{k\in L}\|\min\{x^k,\nabla\hat{f}_{N_{\max}}(x^k,\mu_k)\}\|=\|\min\{x^*,\lim_{k\in L}\nabla\hat{f}_{N_{\max}}(x^k,\mu_k)\}\|.$$

Furthermore, since $\lim_{k\in L}\mu_k = \lim_{k\to\infty}\mu_k = 0$, we obtain

$$\|\min\{x^*, V\}\| = 0, \tag{30}$$

where

$$V = \lim_{x^k\to x^*,\ \mu_k\to 0} \nabla\hat{f}_{N_{\max}}(x^k, \mu_k)$$

and therefore $V \in G_{\hat{f}_{N_{\max}}}(x^*)$.

Since $x^* \geq 0$, (30) implies that $V \geq 0$ and $\min\{x_i^*, V_i\} = 0$ for every $i \in \{1, 2, \ldots, n\}$. Therefore we have that $x_i^* = 0$ or $V_i = 0$. Define $V^+ = \{i \in \{1, 2, \ldots, n\} : V_i > 0\}$ and $V^0 = \{i \in \{1, 2, \ldots, n\} : V_i = 0\}$ and notice that $x_i^* = 0$ for every $i \in V^+$. Let x be an arbitrary nonnegative vector. Then

$$V^T(x-x^*) = \sum_{i=1}^{n} V_i(x_i - x_i^*) = \sum_{i\in V^0} V_i(x_i - x_i^*) + \sum_{i\in V^+} V_i(x_i - x_i^*) = \sum_{i\in V^+} V_i x_i \geq 0.$$

Thus, (30) implies that $V^T(x - x^*) \geq 0$ for every $x \geq 0$ which completes the proof. □

The consequence of the previous lemma and Theorem 2 is that every accumulation point of the sequence $\{x^k\}_{k\in S}$ with S defined by (25) is a Clarke stationary point of function $\hat{f}_{N_{\max}}(x)$. The stronger result is obtained in the following theorem.

Theorem 3. *Suppose that the assumptions of Theorem 1 are satisfied. Then every accumulation point of the sequence $\{x^k\}_{k\in\mathbb{N}}$ is a Clarke stationary point of $\hat{f}_{N_{\max}}(x)$.*

Proof. Theorem 1 implies that for every $k \geq q$

$$\hat{f}_{N_{\max}}(x^{k+1}, \mu_k) \leq \hat{f}_{N_{\max}}(x^k, \mu_k) + \eta\nu_k(d^k)^T\nabla\hat{f}_{N_{\max}}(x^k, \mu_k) + \varepsilon_k.$$

By the induction argument and nonnegativity of the function $\hat{f}_{N_{\max}}$ we obtain that for every $j \in \mathbb{N}$

$$0 \leq \hat{f}_{N_{\max}}(x^q, \mu_q) + \eta\sum_{i=0}^{j-1}\nu_{q+i}(d^{q+i})^T\nabla\hat{f}_{N_{\max}}(x^{q+i}, \mu_{q+i}) + \sum_{i=0}^{j-1}\varepsilon_{q+i} + \bar{\kappa}(\mu_q - \mu_{q+j})$$

and therefore

$$0 \leq -\eta\sum_{i=0}^{\infty}\nu_{q+i}(d^{q+i})^T\nabla\hat{f}_{N_{\max}}(x^{q+i}, \mu_{q+i}) \leq \hat{f}_{N_{\max}}(x^q, \mu_q) + \varepsilon + \bar{\kappa}\mu_q < \infty$$

The above inequalities imply

$$\lim_{k\to\infty}\nu_k(d^k)^T\nabla\hat{f}_{N_{\max}}(x^k, \mu_k) = 0. \tag{31}$$

By the same reasoning as in the proof of Theorem 2, we conclude that (31) implies $\lim_{k\to\infty}(d^k)^T\nabla\hat{f}_{N_{\max}}(x^k,\mu_k)=0$ and

$$\lim_{k\to\infty}\|\min\{x^k,\nabla\hat{f}_{N_{\max}}(x^k,\mu_k)\}\|=0. \tag{32}$$

Algorithm 1 provides the sequence of feasible points and therefore every accumulation point of the sequence $\{x^k\}_{k\in\mathbb{N}}$ is feasible. Moreover, Theorem 2 implies that $\lim_{k\to\infty}\mu_k=0$. Finally, (32) and Lemma 5 imply that every accumulation point of the sequence $\{x^k\}_{k\in\mathbb{N}}$ is a Clarke stationary point of $\hat{f}_{N_{\max}}(x)$. □

We conclude the convergence analysis by stating the conditions under which every accumulation point of the sequence generated by Algorithm 1 is a global optimum.

Theorem 4. *Suppose that the assumptions of Theorem 1 are satisfied and that the set K defined by (24) is infinite. Then every accumulation point x^* of the sequence $\{x^k\}_{k\in\mathbb{N}}$ satisfies $x^*\geq 0$ and*

$$\hat{f}_{N_{\max}}(x^*)=0.$$

Proof. The definition of the search direction implies that $x^k\geq 0$ for every $k\in\mathbb{N}$ and therefore every accumulation point of the sequence $\{x^k\}_{k\in\mathbb{N}}$ is also nonnegative. Let us first prove that

$$\lim_{k\in K}\hat{f}_{N_{\max}}(x^k)=0. \tag{33}$$

Denote $\{x^{k_j}\}_{j\in\mathbb{N}}=\{x^k\}_{k\in K}$. Then Algorithm 3 implies that for every j

$$\hat{f}_{N_{max}}(x^{k_j})\leq\max\{\bar{\xi}\hat{f}_{N_{\max}}(x^{k_j-1}),\frac{|\hat{f}_{N_{\max}}(x^{k_j})-\hat{f}_{N_{\max}}(x^{k_j},\mu_{k_j-1})|}{\alpha}\}. \tag{34}$$

Furthermore, Lemma 4 implies that $|\hat{f}_{N_{\max}}(x^{k_j})-\hat{f}_{N_{\max}}(x^{k_j},\mu_{k_j-1})|\leq\hat{\kappa}\mu_{k_j-1}$. Also, we know that $\{\mu_k\}_{k\in\mathbb{N}}$ is nonincreasing and Algorithm 3 implies

$$\mu_{k_j-1}\leq\mu_{k_j-1}\leq\frac{\alpha}{2\hat{\kappa}}\hat{f}_{N_{\max}}(x^{k_j-1}).$$

Therefore, from (34) we obtain that for every $j\in\mathbb{N}$

$$\hat{f}_{N_{max}}(x^{k_j})\leq r\hat{f}_{N_{\max}}(x^{k_j-1})$$

where $r=\max\{\bar{\xi},1/2\}\in(0,1)$. Since $\hat{f}_{N_{max}}$ is nonnegative, this obviously implies (33).

Since d^k is a descent search direction, the line search implies that for every $k\geq q$

$$\hat{f}_{N_{max}}(x^{k+1},\mu_k)\leq\hat{f}_{N_{max}}(x^k,\mu_k)+\varepsilon_k.$$

Furthermore, Lemma 4 implies that for every j

$$\hat{f}_{N_{max}}(x,\mu_{k+j})\leq\hat{f}_{N_{max}}(x,\mu_k)+\bar{\kappa}(\mu_k-\mu_{k+j}).$$

Using the previous two inequalities and the induction argument we obtain that for every $k, j \in \mathbb{N},\ k \geq q$

$$\hat{f}_{N_{max}}(x^{k+j}, \mu_{k+j}) \leq \hat{f}_{N_{max}}(x^k, \mu_k) + \bar{\kappa}(\mu_k - \mu_{k+j}) + \sum_{i=0}^{j-1} \varepsilon_{k+i}. \qquad (35)$$

Now, let x^* be an arbitrary accumulation point of the sequence $\{x^k\}_{k\in\mathbb{N}}$

$$\lim_{k\in L} x^k = x^*.$$

Since both L and K are infinite, for every $k \in L$ there exists a finite integer $j(k) \geq 0$ such that $k - j(k) = s(k) \in K$. Therefore, inequality (35) and Lemma 4 imply that for every $k \in L$

$$\begin{aligned}
\hat{f}_{N_{max}}(x^k, \mu_k) &= \hat{f}_{N_{max}}(x^{s(k)+j(k)}, \mu_{s(k)+j(k)}) \\
&\leq \hat{f}_{N_{max}}(x^{s(k)}, \mu_{s(k)}) + \bar{\kappa}(\mu_{s(k)} - \mu_{s(k)+j(k)}) + \sum_{i=0}^{j(k)-1} \varepsilon_{s(k)+i} \\
&\leq \hat{f}_{N_{max}}(x^{s(k)}) + \bar{\kappa}(2\mu_{s(k)} - \mu_{s(k)+j(k)}) + \sum_{i=0}^{j(k)-1} \varepsilon_{s(k)+i}.
\end{aligned}$$

Now, the summability of $\{\varepsilon_k\}_{k\in\mathbb{N}}$ implies that $\lim_{k\to\infty} \varepsilon_k = 0$ and since $j(k)$ is finite, we obtain $\lim_{k\in L} \sum_{i=0}^{j(k)-1} \varepsilon_{s(k)+i} = 0$. Moreover, Theorem 2 implies $\lim_{k\to\infty} \mu_k = 0$ and

$$0 \leq \hat{f}_{N_{max}}(x^*) = \lim_{k\in L} \hat{f}_{N_{max}}(x^k, \mu_k) \leq \lim_{k\in L} \hat{f}_{N_{max}}(x^{s(k)}) = \lim_{j\in K} \hat{f}_{N_{max}}(x^j) = 0$$

which completes the proof. □

5 Numerical results

In this section we present some numerical results obtained by applying two algorithms on the set of test problems which can be found in Chen et al. [5] and Li et al. [9]. Our aim is to compare the performance of Algorithm 1, which we refer to as VSS, with the results obtained by algorithm proposed in Li et al. [9] which we call LLS. The key differences between these two methods lays in the fact that VSS uses the variable sample size scheme. Also, the line search in VSS is nonmonotone and the updating of the smoothing parameters is more complex than in LLS.

In order to provide a better insight into the results, we state the relevant notation considering the test problems. The point $\hat{x}$ has exactly n_x positive components which are chosen randomly from $(0, \tau)$ where $\tau = 10^{-6}$. Interval $(-\sigma, \sigma)$ represents the range of elements of $E(M(\omega)) - M(\omega^j)$ for $j = 1, 2, \ldots, N_{max}$. On the other hand, $[0, \beta_e)$ is the range of elements of $(M(\omega^j)\hat{x} + q(\omega^j))_i$ for all i

such that $\hat{x}_i > 0$. Parameter β_e is especially important since $\beta_e = 0$ implies that the point $\hat{x}$ is the unique solution of the considered problem and the optimal value of the objective function is 0.

The initial points are set to $x^0 = \lfloor v + 10u \rfloor$ where v is the vector with all components equal to 1 and u is a random vector from uniform distribution. The stopping criterion for both algorithms is

$$\| \min\{x^k, \nabla \hat{f}_{N_{max}}(x^k, \mu_k)\} \| \leq \gamma\epsilon \quad \text{and} \quad \mu_k \leq \epsilon,$$

with $\gamma = 100$ and $\epsilon = 10^{-6}$. The search direction d^k is the BB direction obtained with $\alpha_{min} = \alpha_0 = 0.1$. The line search is performed with $\beta = \eta = 0.5$. The initial smoothing parameter is $\mu_0 = 1$. In the LLS method, the smoothing parameter is updated as $\mu_k = 0.5\mu_{k-1}$, while the parameters of the Algorithm 3 are $\alpha = \bar{\xi} = 0.5$ and $\hat{\kappa} = 1$ and in step S2 we set

$$\mu_{k+1} = \min\{\epsilon, \frac{\mu_k}{2}, \frac{\alpha\tilde{\beta}}{2\bar{\kappa}}, \bar{\bar{\mu}}(x^{k+1}, \gamma\tilde{\beta})\}.$$

The sequence that makes the line search in VSS nonmonotone is initialized by $\varepsilon_0 = \max\{1, \hat{f}_{N_0}(x^0, \mu_0)\}$ and it is updated only if the sample size does not change. More precisely, if $N_{k-1} = N_k$ we set $\varepsilon_k = \varepsilon_0 k^{-1.1}$. Otherwise, $\varepsilon_k = \varepsilon_{k-1}$. In VSS we set $N_0 = N_0^{min} = 3$, while the rest of the parameters for updating the sample size are $d = 0.5$, $\delta = 0.95$ and $\tilde{\nu_1} = 1/\sqrt{N_{max}}$.

For each test problem we conducted 10 different runs of both algorithms. The results in the following two tables represent the average values of successful runs reported in column s. A run is considered successful if the number of function evaluations (fev) needed to satisfy the stopping criterion does not exceed 10^7. The number of function evaluations counts the evaluations of the function F and the gradient $\nabla_x F$, and each component of the gradient is counted as one function evaluation. The column *stac* refers to the measure of stationarity $\| \min\{x^k, \nabla \hat{f}_{N_{max}}(x^k, \mu_k)\} \|$.

As one of the referees pointed out the algorithm would reduce to the SAA method if $N_k = N_{\max},\ k = 1, 2, \ldots$ and the whole procedure would not yield any gain in the terms of computational costs. We observed the following average values for the sample size in the tested examples. The iteration number in which $N_{\max}$ was taken for the first time varies from $k = 10$ to $k = 60$ with the average being $\bar{k} = 24$ among all runs. The iteration number after which the sample size stays constant i.e. $N_k = N_{max}$ for $k \geq q$ varies between $q = 31$ and $q = 170$ with the average $\bar{q} = 78$ among all runs. The total number of iterations needed to satisfy the exit criterion changes from $k_f = 59$ to $k_f = 186$ with the average $\bar{k}_f = 98$. Speaking in relative terms over all runs the maximal sample size is taken after 25.60% of iterations needed and the sample size becomes stationary with $N_k = N_{\max}$ in the last 23.76% iterations. Thus the behavior of the sample size is as expected and yields a significant reduction in computational costs.

Another key moment in the efficiency of VSS method is a good balance of N_k and μ_k. Instead of decreasing μ_k in each iteration, we apply the updating procedure that is closely connected with the rule for N_k. Both N_k and μ_k clearly

influence the precision of the approximate objective function and iteration cost. The procedures stated in Algorithms 2 and 3 ensure that this fact is taken into account. During the initial stage, while $N_k < N_{\max}$, the new sample size N_{k+1} depends on μ_k as $\varepsilon_\delta^N = \varepsilon_\delta^N(x^k, \mu_k)$. If the approximate objective function is taken with an increased precision i.e. if $N_{k+1} > N_k$ then the smoothing parameter is also decreased but otherwise it is kept fixed. Finally, when $N_{\max}$ is reached, the update for μ_k given in S2 of Algorithm 3 takes into account the objective function value, the precision of the previous smoothing approximation $\hat{f}_{N_{\max}}(x^{k+1}, \mu_k)$ as well as the threshold value. Notice that the stationarity criterion in S3 of Algorithm 1 is defined by μ_k and thus μ_k fundamentally influences the scheduling sequence. The numerical test we performed indicate that this condition in S3 is fundamental for the efficiency of Algorithm 1 as it prevents unproductive steps with smaller sample sizes $N_k < N_{\max}$.

$\beta_e = 0$	VSS			LLS		
$(N_{max}, n, n_x, \sigma)$	$stac$	fev	s	$stac$	fev	s
(100,20,10,20)	4.5672E-05	6.2264E+04	10	5.0399E-05	1.4733E+05	10
(100,20,10,10)	7.6566E-05	6.7011E+04	10	6.7901E-05	1.3712E+05	10
(100,20,10,0)	9.4806E-05	2.2688E+06	2	9.0334E-05	8.6548E+05	10
(100,40,20,20)	6.4237E-05	1.2939E+05	10	5.3148E-05	2.7244E+05	10
(100,40,20,10)	5.4754E-05	1.2830E+05	10	4.4507E-05	2.6979E+05	10
(100,40,20,0)	-	-	0	8.9627E-05	2.5557E+06	10
(200,60,30,20)	4.6531E-05	3.7331E+05	10	4.3733E-05	7.9708E+05	10
(200,60,30,10)	6.1972E-05	3.4602E+05	8	5.0888E-05	8.1134E+05	10
(200,60,30,0)	-	-	0	9.1281E-05	7.2546E+06	9
(200,80,40,20)	6.0005E-05	5.0364E+05	9	6.0231E-05	1.0715E+06	10
(200,80,40,10)	4.8242E-05	4.8897E+05	8	5.7896E-05	1.0585E+06	10
(200,80,40,0)	-	-	0	9.0104E-05	8.4345E+06	6
(200,100,50,20)	5.2790E-05	6.3805E+05	10	3.7452E-05	1.2967E+06	10
(200,100,50,10)	4.3754E-05	6.2113E+05	8	5.3203E-05	1.2577E+06	10
(200,100,50,0)	-	-	0	9.2829E-05	9.9740E+06	2
(300,120,60,20)	6.5184E-05	1.2418E+06	8	5.9607E-05	2.4488E+06	10
(300,120,60,10)	4.4089E-05	1.4401E+06	8	5.1507E-05	2.3362E+06	10
(300,120,60,0)	-	-	0	-	-	0
(1000,50,25,10)	7.0037E-05	1.3376E+06	4	4.4642E-05	3.0532E+06	10
(1000,50,25,0)	-	-	0	-	-	0

Table 1: VSS versus LLS, $\beta_e = 0$

$\beta_e > 0$	VSS			LLS		
$(N_{max}, n, n_x, \sigma, \beta_e)$	$stac$	fev	s	$stac$	fev	s
(100,20,10,20,10)	3.5346E-05	5.7558E+04	10	6.3513E-05	1.3971E+05	10
(100,20,10,20,5)	3.2922E-05	5.6040E+04	10	5.0903E-05	1.3866E+05	10
(100,40,20,20,10)	6.5369E-05	1.1719E+05	10	5.8538E-05	2.6977E+05	10
(100,40,20,20,5)	5.6539E-05	1.4118E+05	10	5.4201E-05	2.6914E+05	10
(100,40,20,10,20)	6.1957E-05	1.5586E+05	10	5.3656E-05	2.9287E+05	10
(200,60,30,20,10)	4.0542E-05	3.3951E+05	10	4.8921E-05	8.0256E+05	10
(200,60,30,20,5)	4.9567E-05	2.7194E+05	10	6.2701E-05	7.5920E+05	10
(200,80,40,20,10)	3.8306E-05	4.8032E+05	10	7.2183E-05	1.0549E+06	10
(200,80,40,20,5)	5.0707E-05	4.0370E+05	10	6.9662E-05	1.0168E+06	10
(200,100,50,20,10)	4.6138E-05	5.8055E+05	10	7.1681E-05	3.3134E+06	10
(200,100,50,20,5)	4.6833E-05	5.5841E+05	10	6.1738E-05	1.2628E+06	10
(200,100,50,10,20)	4.9498E-05	8.0825E+05	10	5.4844E-05	1.5759E+06	10
(300,120,60,20,10)	6.2990E-05	1.0970E+06	10	5.8830E-05	2.3564E+06	10
(300,120,60,20,5)	5.2634E-05	8.6039E+05	10	7.5739E-05	2.2575E+06	10
(300,120,60,10,20)	5.4974E-05	1.4747E+06	10	6.5885E-05	4.2844E+06	10
(1000,50,25,20,10)	3.6456E-05	1.3141E+06	10	4.2850E-05	3.2556E+06	10
(1000,50,25,10,5)	3.3576E-05	8.4020E+05	10	4.9866E-05	2.9910E+06	10
(1000,100,50,5,10)	4.1478E-05	2.0831E+06	10	5.3271E-05	5.8908E+06	10
(1000,100,50,10,5)	4.1682E-05	1.7271E+06	10	5.8140E-05	6.1028E+06	10

Table 2: VSS versus LLS, $\beta_e > 0$

Table 1 contains the results obtained by considering the test problems with $\beta_e = 0$, while Table 2 states the results for $\beta_e > 0$. Notice that in the latter case all of the runs were successful, while $\beta_e = 0$ caused failure in many tested problems, especially regarding VSS algorithm. The examples with $\sigma = 0$ turn out to be the most challenging for VSS, but these particular problem settings also affected LLS performance. Although the algorithm LLS seems to be more stable, VSS gains the advantage in the fev column and therefore the performance of the tested algorithms is comparable. The results in Table 2 reveal the clear advantage of VSS method if $\beta_e > 0$. In all the tested problems the average number of function evaluations for the VSS is lower than fev for LLS. Therefore, our conclusion is that the overall results suggest that variable sample scheme with smoothing proposed in this paper reduces the computational cost of the SAA approach.

Acknowledgment We are grateful to the anonymous referees whose constructive remarks helped us to improve this paper.

References

[1] J. Barzilai and J.M. Borwein, *Two-point step size gradient methods*, IMA J. Numer. Anal. 8 (1988), pp. 141-148.

[2] C. Chen and O.L. Mangasarian, *A class of smoothing functions for linear and mixed complementarity problems*, Comput. Optim. Appl. 5 (1996), pp. 97-138.

[3] X. Chen and M. Fukushima, *Expected residual minimization method for stochastic linear complementarity problems*, Math. Oper. Res. 30 (2005), pp. 1022-1038.

[4] X. Chen, L. Qi and D. Sun, *Global and superlinear convergence of the smoothing Newton method and its application to general boxed constrained variational inequalities*, Math. Comp. 67 (1998), pp. 519-540.

[5] X. Chen, C. Zhang and M. Fukushima, *Robust solution of monotone stochastic linear complementarity problems*, Springer, Math. Program. 117 (2009), pp. 51-80.

[6] F.H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley, New York, 1983.

[7] W. Cottle, J.S. Pang and R.E. Stone, *The Linear Complementarity Problem*, Academic Press, Boston, 1992.

[8] W. Hager, B.A. Mair and H. Zhang, *An affine scaling interior-point CBB method for box-constrained optimization*, Math. Program. 119 (2009), pp. 1-32.

[9] X. Li, H. Liu and X. Sun, *Feasible smooth method based on Barzilai-Borwein method for stochastic linear complementarity problem*, Numer. Algorithms 57 (2011), pp. 207-215.

[10] C. Kanzow and H. Pieper, *Jacobian smoothing methods for general nonlinear complementarity problems*, SIAM J. Optim. 9 (1999), pp. 342-373.

[11] N. Krejić and N. Krklec, *Line search methods with variable sample size for unconstrained optimization*, J. Comput. Appl. Math. 245 (2013), pp. 213-231.

[12] N. Krejić and N. Krklec Jerinkić, *Nonmonotone line search methods with variable sample size*, Numerical Algorithms (2014) doi 10.1007/s11075-014-9869-1.

[13] N. Krejić and S. Rapajić, *Globally convergent Jacobian smoothing inexact Newton methods for NCP*, Comput. Optim. Appl. 41 (2008), pp. 243-261.

[14] M. Raydan, *The Barzilai and Borwein Gradient Method for the Large-Scale Unconstrained Minimization Problem*, SIAM J. Optim. 7 (1997), pp. 26-33.

[15] C. Zhang and X. Chen, *Smoothing projected gradient method and its application to stochastic linear complementarity problems*, SIAM J. Optim. 20 (2009), pp. 627-649.