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

Nataša Krejić, Nataša Krklec Jerinkić, Sanja Rapajić
Department of Mathematics and Informatics, University of Novi Sad,
Trg Dositeja Obradovića 4, 21000 Novi Sad, Serbia

October 5, 2013

Abstract. ABSTRACT

Key words. Stochastic linear complementarity problems, variable sample size, semismooth systems, BB direction

AMS subject classifications. 65H10 ; 90C33

1 Introduction

Uvodna prica!!!!!!!!!!!!!!!!!!!!!!!!!

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

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

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

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

where $\Phi_\omega : R^n \times \Omega \to R^n$ is defined by

$$\Phi_\omega(x) = \begin{bmatrix} \phi\left(x_1, [M(\omega)x]_1 + q_1(\omega)\right) \\ \phi(x_2, [M(\omega)x]_2 + q_2(\omega)) \\ \vdots \\ \phi(x_n, [M(\omega)x]_n + q_n(\omega)) \end{bmatrix},$$

and $\phi : R^2 \to R$ is a NCP function. In this paper, we focus on the ERM reformulation based on "min" function $\phi(a, b) = \min\{a, b\}$ which is expressed as

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

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 $\Omega = \{\omega^1, \omega^2, \ldots, \omega^N\}$ is a finite set.

Since we focus on

$$\Phi_\omega(x) = \min\{x, M(\omega)x + q(\omega)\} \tag{2}$$

and if we define $\Theta_\omega(x) = \frac{1}{2}\|\Phi_\omega(x)\|^2$ then (1) is equivalent to

$$f(x) = E(2\Theta_\omega(x)) \to \min, \quad x \geq 0. \tag{3}$$

The function $f(x)$ in ERM reformulation is in the form of mathematical expectation, so it is well known that in general, is very difficult to compute it accurately. Because of that, the Sample Average Approximation (SAA) is usually employed in practice for estimating $f(x)$.

Assume that $\{\omega^1, \omega^2, ..., \omega^N\}$ from Ω is a sample of random vectors that are independent and identically distributed. Then problem (1) can be approximated by Monte Carlo sampling with

$$\hat{f}_N(x) = \frac{1}{N}\sum_{j=1}^{N} F(x, \omega^j) \to \min, \quad x \geq 0, \tag{4}$$

where $F(x, \omega^j) = 2\Theta_{\omega^j}(x) = \|\Phi_{\omega^j}(x)\|^2$ and our focus will be on it. Since the problem (4) is nonsmooth, it can be solved by using smoothing methods. In this paper we propose a new smoothing method based on Brazilai-Borwein search direction defined in Li et al. [9].

The corresponding smoothing problem for ERM reformulation was introduced in Li et al. [9] and Zhang, Chen [14] and is defined for a smoothing parameter $\mu > 0$.

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

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

A smoothing function for $\Theta_\omega(x)$ is

$$\tilde{\Theta}_\omega(x, \mu) = \frac{1}{2}\|\tilde{\Phi}_\omega(x, \mu)\|^2 = \frac{1}{2}\left\| \begin{bmatrix} \phi\left(x_1, [M(\omega)x]_1 + q_1(\omega), \mu\right) \\ \phi(x_2, [M(\omega)x]_2 + q_2(\omega), \mu) \\ \vdots \\ \phi(x_n, [M(\omega)x]_n + q_n(\omega), \mu) \end{bmatrix} \right\|^2,$$

where $\tilde{\Phi}_\omega(x, \mu)$ is a smoothing function for $\Phi_\omega(x)$. Its components

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

are defined with (5). The smoothing approximation for $f(x)$ is

$$\tilde{f}(x, \mu) = E(2\tilde{\Theta}_\omega(x, \mu)). \tag{7}$$

Using SAA method, function $\tilde{f}(x, \mu)$ can be estimated with

$$\hat{f}_N(x, \mu) = \frac{2}{N}\sum_{j=1}^{N} \tilde{\Theta}_{\omega^j}(x, \mu), \tag{8}$$

where $\{\omega^1, \omega^2, ..., \omega^N\}$ from Ω is a sample realization generated at the beginning of optimization process.

The idea is to consider the smoothing functions

$$\hat{f}_N(x, \mu) = \frac{1}{N} \sum_{j=1}^{N} F(x, \omega^j, \mu),$$

$\mu > 0$, where $F(x, \omega^j, \mu) = 2\tilde{\Theta}_{\omega^j}(x, \mu) = \|\tilde{\Phi}_{\omega^j}(x, \mu)\|^2$, instead of the objective function $\hat{f}_N(x)$ from (4).

More precisely, at each iteration of algorithm, the objective function $\hat{f}_N(x)$ can be approximated by a smooth function $\hat{f}_N(x, \mu_k)$ with a fixed smoothing parameter $\mu_k > 0$. The sequence of smoothing functions $\hat{f}_N(x, \mu_k)$ will tend to nonsmooth objective function $\hat{f}_N(x)$ when $\mu_k \to 0$.

The disadvantage of SAA method is using large sample size N in every iteration, which makes very expensive calculating the objective function. Because of that variable sample size strategies are recommended. In order to make the process significantly cheaper, in our algorithm we use the line search strategy with variable sample size proposed in Krejić, Krklec [11], [12].

So, except the smoothing parameter μ_k, every iteration has its own sample size N_k. Therefore, at k-th iteration we consider the function

$$\hat{f}_{N_k}(x, \mu_k) = \frac{1}{N_k} \sum_{j=1}^{N} F(x, \omega^j, \mu_k). \tag{9}$$

This function is differentiable with the gradient

$$\nabla \hat{f}_{N_k}(x, \mu_k) = \frac{1}{N_k} \sum_{j=1}^{N_k} \nabla F(x, \omega^j, \mu_k),$$

where $\nabla F(x, \omega^j, \mu_k) = 2\tilde{\Phi}'_{\omega^j}(x, \mu_k)^T \tilde{\Phi}_{\omega^j}(x, \mu_k)$ and $\tilde{\Phi}'_{\omega^j}$ is the Jacobian of function $\tilde{\Phi}_{\omega^j}$.

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

2 Preliminaries

A few words about notation and definitions. Throughout the paper $\|\cdot\|$ represents the Euclidian norm, $\|\cdot\|_F$ the Frobenius 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 and satisfy

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

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 with $[\nabla g(x)]_i$. 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 denoted by $dist(A, \mathcal{A}) = \inf_{B\in\mathcal{A}} \|A - B\|$, 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

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

Let consider the problem of the form

$$g(x) \to \min, \quad x \geq 0, \tag{10}$$

where $g : R^n \to R$.

In Zhang, Chen [14] authors have pointed out that if g is differentiable at x^*, then x^* is a local minimizer of g if and only if $\|\min\{x^*, \nabla g(x^*)\}\| = 0$.

Note that $\|\min\{x^*, \nabla g(x^*)\}\| = 0$ if and only if x^* is a stationary point of g, that is if

$$\langle \nabla g(x^*), x - x^* \rangle \geq 0, \tag{11}$$

holds for any $x \geq 0$.

If the function g from (10) is a local Lipschitz continuous function, then 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 the convex hull and D_g be the subset of R^n where g is differentiable.

Following definitions are related with this locally Lipschitzian function g from (10).

Definition 1 *[14] Function $\tilde{g} : R^n \times R_+ \to R$ is a smoothing function of 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.

Definition 2 *[14] x^* is a Clarke stationary point of $g(x)$, if there exists $V \in \partial g(x^*)$ such that*

$$\langle V, x - x^* \rangle \geq 0,$$

for every $x \geq 0$.

Definition 3 *[14] For any fixed $x^* \geq 0$ let us denote*

$$G_{\tilde{g}}(x^*) = \{V, \lim_{x^k \to x^*, \mu_k \to 0} \nabla \tilde{g}(x^k, \mu_k) = V\}.$$

It is shown in Zhang, Chen [14] that $G_{\tilde{g}}(x^*)$ is nonempty and bounded set and $G_{\tilde{g}}(x^*) \subseteq \partial g(x^*)$, for any $x^* \in R^n$.

Definition 4 *[14] x^* is a stationary point of function $g(x)$ associated with a smoothing function $\tilde{g}(x, \mu)$, if there exists $V \in G_{\tilde{g}}(x^*)$ such that*

$$\langle V, x - x^* \rangle \geq 0,$$

for any $x \geq 0$.

3 The algorithm

As we have mentioned before, the ERM problem (1) can be approximated with nonsmooth problem (4) and our focus is on solving this problem by using smoothing technique. In every iteration we approximate the nonsmooth objective function with the smoothing function $\hat{f}_{N_k}(x, \mu_k)$ defined by (9) and

try to solve this smoothing optimization problem by using Barzilai-Borwein gradient method.

!!! Reci zasto BB pravac i reci da je za njih zgodno koristiti nemonotone line-search tehnike iz tog i tog razloga. Reci zasto bas nonmonotone line search with variable sample size. Zato je u algoritmu koricsen nonmonotone line search technique with variable sample size, presented in Krejić, Krklec [12].

So, the search direction which is used in our algorithm is smooth Barzilai-Borwein (BB) direction defined in Li et al. [9] in this way

$$d_i^k = \begin{cases} -\frac{1}{\alpha_k}[\nabla \hat{f}_{N_k}(x^k, \mu_k)]_i, & \text{if} \quad i \in I_1(x^k) \\ -\frac{[\nabla \hat{f}_{N_k}(x^k, \mu_k)]_i}{\alpha_k + \frac{[\nabla \hat{f}_{N_k}(x^k, \mu_k)]_i}{x_i^k}}, & \text{if} \quad i \in I_2(x^k) \\ -x_i^k, & \text{if} \quad i \in I_3(x^k) \end{cases}, \tag{12}$$

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

$$I_1(x^k) = \{i, \;\; i \in \{1, ..., n\}, \quad [\nabla \hat{f}_{N_k}(x^k, \mu_k)]_i \leq 0\},$$

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

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

and

$$\alpha_k = \begin{cases} \alpha_k^{BB1}, & \text{if} \quad mod(k, 4) = 0, 1 \\ \alpha_k^{BB2}, & \text{if} \quad mod(k, 4) = 2, 3 \end{cases}, \tag{13}$$

$$\alpha_k^{BB1} = \begin{cases} \max\{\alpha_{\min}, \frac{(s^{k-1})^T y^{k-1}}{\|s^{k-1}\|^2}\}, & \text{if} \quad \|s^{k-1}\| > 0 \\ \alpha_{\min}, & \text{else} \end{cases},$$

$$\alpha_k^{BB2} = \begin{cases} \max\{\alpha_{\min}, \frac{\|y^{k-1}\|^2}{(s^{k-1})^T y^{k-1}}\}, & \text{if} \quad (s^{k-1})^T y^{k-1} \neq 0 \\ \alpha_{\min}, & \text{else} \end{cases},$$

$$\alpha_{\min} > 0, \quad s^{k-1} = x^k - x^{k-1}, \quad y^{k-1} = \nabla \hat{f}_{N_k}(x^k, \mu_k) - \nabla \hat{f}_{N_{k-1}}(x^{k-1}, \mu_{k-1}).$$

We can now state the main algorithm as follows.

ALGORITHM 1

S0 Input parameters: $N_{\max}, N_0^{\min} \in N,\ x^0 \in R_+^n,\ \bar{\kappa}, \gamma > 0,\ 0 < \alpha_{\min} \leq \alpha_0 < 1,\ \delta, \eta, \beta, \tau, \tilde{\nu}_1, \bar{\xi}, \alpha, d \in (0,1)$. Let $\{\varepsilon_k\}_{k \in N}$, be a sequence such that $\varepsilon_k > 0$, $\sum_{k \in N} \varepsilon_k \leq \varepsilon < \infty$.

S1 Generate the sample realization: $\omega^1, \ldots, \omega^{N_{\max}}$.

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 $\hat{f}_{N_k}(x^k, \mu_k)$, $\varepsilon_\delta^{N_k}(x^k, \mu_k)$ and $\nabla \hat{f}_{N_k}(x^k, \mu_k)$.

S3 If $\| \min\{x^k, \nabla \hat{f}_{N_k}(x^k, \mu_k)\}\| = 0$

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

2) if $N_k < N_{\max}$ and $\varepsilon_\delta^{N_k}(x^k, \mu_k) = 0$ 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 S12.
If $\| \min\{x^k, \nabla \hat{f}_{N_k}(x^k, \mu_k)\}\| > 0$, go to step S4.

S4 Determine the BB direction d^k by (12) and (13).

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 the candidate sample size N_k^+ using Algorithm 2.

S8 Determine the sample size N_{k+1}.

S9 Determine μ_{k+1} using Algorithm 3 and then the lower bound $N_{k+1}^{\min}$.

S10 Compute $y^k = \nabla \hat{f}_{N_{k+1}}(x^{k+1}, \mu_{k+1}) - \nabla \hat{f}_{N_k}(x^k, \mu_k)$.

S11 Determine α_{k+1} by (13).

S12 Set $k = k + 1$ and go to step S2.

Updating the sample size refers to the algorithms stated in [12], while the slightly different versions of these algorithms can be found in [11]. 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

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

where

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

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

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

where $d \in (0, 1]$ and $N_k^{\min} \leq N_k^+ \leq N_{\max}$. 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 we decrease the sample size if possible. We state the algorithm for choosing the candidate sample size N_k^+.

ALGORITHM 2

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

S1 Determine N_k^+

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

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

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

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

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

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 $\rho_k \geq \frac{N_k - N_k^+}{N_k}$ we set $N_{k+1} = N_k$. In all the other cases, the decrease is accepted and $N_{k+1} = N_k^+$. Notice that $N_{k+1} \geq N_k^+$ 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

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

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 of updating the smoothing parameter, where $\bar{\bar{\mu}}(x^{k+1}, \gamma\tilde{\beta})$ is a threshold value for the smoothing parameter, which will be defined later in the next section.

ALGORITHM 3

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

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

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

else $\mu_{k+1} = \mu_k$.

It is easy to see that BB direction d^k defined by (12) is feasible and descent search direction for $\hat{f}_{N_k}(x^k, \mu_k)$, because $x^k + d^k \geq 0$ and $\nabla\hat{f}_{N_k}(x^k, \mu_k)^T d^k < 0$ if $\| \min\{x^k, \nabla\hat{f}_{N_k}(x^k, \mu_k)\}\| \neq 0$.

4 Convergence analysis

Our algorithm is based on the idea of using line search with variable sample size presented in Krejić, Krklec [12], so the following assumptions are necessary for applying that idea:

A1 : For every ω and $\mu \in R_{++}$, $F(\cdot, \omega, \mu) \in \mathcal{C}^1(R^n)$,

A2 : There exist finite constants M_0 such that $M_0 \leq F(x, \omega, \mu)$, for every ω, x, μ.

Let $\Phi_\omega(x)$, $f(x)$, $\hat{f}_N(x)$, $\tilde{\Phi}_\omega(x, \mu)$, $\tilde{f}(x, \mu)$ and $\hat{f}_N(x, \mu)$ be the functions defined by (2)- (8) respectively and $\Omega = \{\omega^1, \ldots, \omega^N\}$. First we will give some properties of these functions which are necessary for convergence analysis.

Since $F(x, \omega^j, \mu) = \|\tilde{\Phi}_{\omega^j}(x, \mu)\|^2$, it is easy to see that $F(\cdot, \omega^j, \mu) \in \mathcal{C}^1(R^n)$ and $F(x, \omega^j, \mu) \geq 0$, for every $\omega^j \in \Omega$, $\mu \in R_{++}$, $x \in R^n$, which means that the assumptions A1 and A2 are satisfied and also imply that $\hat{f}_N(\cdot, \mu) \in \mathcal{C}^1(R^n)$ and

$$\hat{f}_N(x, \mu) \geq 0, \tag{14}$$

for every $x \in R^n$ and every $n \in N$.

Lemma 1 *[14] 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 function $\tilde{\Phi}_\omega(x,\mu)$ has the Jacobian consistency property defined in Chen et al. [4] which means

$$\lim_{\mu\to 0} \tilde{\Phi}'_\omega(x,\mu) \in \partial_C\Phi_\omega(x). \tag{15}$$

It is proved in Zhang, Chen [14] that $\tilde{f}(x,\mu)$ given in (7) is a smoothing function for $f(x)$ given in (3), because it satisfies Definition 1.

The sample realization generated before the optimization process in our algorithm is $\Omega = \{\omega^1, \ldots, \omega^{N_{\max}}\}$ and from now on we consider this Ω.

In order to determine the threshold value we need following lemmas.

Lemma 2 *Let $x \in R^n$. Then*

$$\lim_{\mu\to 0} \nabla\hat{f}_{N_{\max}}(x,\mu) \in \partial\hat{f}_{N_{\max}}(x).$$

Proof. Since, by Lemma 1, smoothing function $\tilde{\Phi}_\omega(x,\mu)$ has the Jacobian consistency property (15), it means that

$$\lim_{\mu\to 0} \tilde{\Phi}'_{\omega^j}(x,\mu) \in \partial_C\Phi_{\omega^j}(x), \tag{16}$$

for every ω^j, $j = 1, ..., N_{\max}$. From (16) and the fact that $\tilde{\Phi}_{\omega^j}(x,\mu)$ is the smoothing function for $\Phi_{\omega^j}(x)$ there follows

$$\lim_{\mu\to 0} \tilde{\Phi}'_{\omega^j}(x,\mu)^T\tilde{\Phi}_{\omega^j}(x,\mu) \in \partial_C\Phi^T_{\omega^j}(x)\Phi_{\omega^j}(x), \tag{17}$$

for every ω^j, $j = 1, ..., N_{\max}$. Since $\nabla\hat{f}_{N_{\max}}(x,\mu) = \frac{2}{N_{\max}} \sum_{j=1}^{N_{\max}} \tilde{\Phi}'_{\omega^j}(x,\mu)^T\tilde{\Phi}_{\omega^j}(x,\mu)$, by (17) and the definition of $\partial\hat{f}_{N_{\max}}(x)$ we have $\lim_{\mu\to 0} \nabla\hat{f}_{N_{\max}}(x,\mu) \in \partial\hat{f}_{N_{\max}}(x)$, which completes the proof. □

An immediate consequence of Lemma 2 is that for every fixed $\delta_1 > 0$, there exists a threshold value $\bar{\bar{\mu}}(x, \delta_1) > 0$ such that

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

for all $0 < \mu \leq \bar{\bar{\mu}}(x, \delta_1)$.

It is important for the algorithm design to have an explicit expression of the threshold value $\bar{\bar{\mu}}(x, \delta_1) > 0$, because updating the smoothing parameter depends on it.

Since (15) holds by Lemma 1, there follows that for every fixed $\delta > 0$ there exists a threshold value $\bar{\mu}(x, \delta) > 0$ such that

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

for every $0 < \mu \leq \bar{\mu}(x, \delta)$.

The following lemma gives a precise definition of the threshold value $\bar{\mu}(x, \delta)$ for the smoothing parameter.

Lemma 3 *Let $x \in R^n$ be arbitrary but fixed and $\omega \in \Omega$ be fixed. Assume that x is not a solution of SLCP. Let us define*

$$\gamma(x) := \max_i \{\|[M(\omega)]_i - e^i\|, i = 1, ..., n\} \geq 0$$

and

$$\xi(x) := \min_{i \notin \beta(x)} \{|x_i - [M(\omega)x]_i - q_i(\omega)|\} > 0,$$

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

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

Then

$$dist_F \left(\tilde{\Phi}'_\omega(x, \mu), \partial_C \Phi_\omega(x)\right) \leq \delta \tag{18}$$

for all μ such that $0 < \mu \leq \bar{\mu}(x, \delta)$.

Proof. It is shown in Lemma 3.2 Kanzow, Pieper [10] that

$$dist_F\left(\tilde{\Phi}'_\omega(x,\mu), \partial_C\Phi_\omega(x)\right) = \sqrt{\sum_{i=1}^{n} dist_2([\tilde{\Phi}'_\omega(x,\mu)]_i, \partial[\Phi_\omega(x)]_i)^2}, \tag{19}$$

so for proving (18), it is sufficient to show

$$dist_2([\tilde{\Phi}'_\omega(x,\mu)]_i, \partial[\Phi_\omega(x)]_i) \leq \frac{\delta}{\sqrt{n}} \tag{20}$$

for every $i = 1, ..., n$, where

$$dist_2([\tilde{\Phi}'_\omega(x,\mu)]_i, \partial[\Phi_\omega(x)]_i) = \|\tilde{[}\Phi'_\omega(x,\mu)]_i - [V]_i\|. \tag{21}$$

$[V]_i$ is the i-th row of matrix $V \in \partial_C\Phi_\omega(x)$, i.e. $[V]_i \in \partial[\Phi_\omega(x)]_i$ and has the form

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

while the i-th row of matrix $\tilde{\Phi}'_\omega(x,\mu)$ has the form

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

where $y_i = \frac{1}{\mu}(x_i - [M(\omega)x]_i - q_i(\omega) + \mu/2)$. Therefore, we distinguish three different cases:
Case 1. If $x_i - [M(\omega)x]_i - q_i(\omega) \geq \mu/2$ then $x_i > [M(\omega)x]_i + q_i(\omega)$, so

$$\|[\tilde{\Phi}'_\omega(x,\mu)]_i - [V]_i\| = \|[M(\omega)]_i - [M(\omega)]_i\| = 0 \leq \frac{\delta}{\sqrt{n}}. \tag{22}$$

Case 2. If $x_i - [M(\omega)x]_i - q_i(\omega) \leq -\mu/2$ then $x_i < [M(\omega)x]_i + q_i(\omega)$, so

$$\|[\tilde{\Phi}'_\omega(x,\mu)]_i - [V]_i\| = \|e^i - e^i\| = 0 \leq \frac{\delta}{\sqrt{n}}. \tag{23}$$

Case 3. If $-\mu/2 < x_i - [M(\omega)x]_i - q_i(\omega) < \mu/2$ then

$[M(\omega)x]_i + q_i(\omega) - \mu/2 < x_i < [M(\omega)x]_i + q_i(\omega) + \mu/2$, so there are three different possibilities in this case:

L1. If $x_i < [M(\omega)x]_i + q_i(\omega)$ then

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

Now, we want to show

$$\frac{1}{\mu}(x_i - [M(\omega)x]_i - q_i(\omega) + \mu/2)\gamma(x) \leq \frac{\delta}{\sqrt{n}} \tag{24}$$

for all $0 < \mu \leq \bar{\mu}(x,\delta)$. If $\gamma(x) = 0$ then (24) holds trivially for every $\mu > 0$ and also for $\mu \leq \bar{\mu}(x,\delta) = 1$. Hence, suppose that $\gamma(x) \neq 0$. If $\frac{1}{2} - \frac{\delta}{\sqrt{n}\gamma(x)} \leq 0$ then (24) holds also for $\mu \leq \bar{\mu}(x,\delta) = 1$. Otherwise, if $\frac{1}{2} - \frac{\delta}{\sqrt{n}\gamma(x)} > 0$ then (24) holds for every $0 < \mu \leq \frac{\xi(x)2\sqrt{n}\gamma(x)}{\sqrt{n}\gamma(x) - 2\delta} := \bar{\mu}(x,\delta)$, and we obtain the upper bound for μ. We proved (24) which implies

$$\|[\tilde{\Phi}'_\omega(x,\mu)]_i - [V]_i\| \leq \frac{\delta}{\sqrt{n}}. \tag{25}$$

L2. If $x_i > [M(\omega)x]_i + q_i(\omega)$ then

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

Now, we want to show

$$(1 - \frac{1}{\mu}(x_i - [M(\omega)x]_i - q_i(\omega) + \mu/2))\gamma(x) \leq \frac{\delta}{\sqrt{n}} \tag{26}$$

for all $0 < \mu \leq \bar{\mu}(x, \delta)$. If $\gamma(x) = 0$ then (26) holds for every $\mu \leq \bar{\mu}(x, \delta) = 1$. Suppose that $\gamma(x) \neq 0$. If $\frac{1}{2} - \frac{\delta}{\sqrt{n}\gamma(x)} \leq 0$ then (26) holds also for $\mu \leq \bar{\mu}(x, \delta) = 1$. Otherwise, if $\frac{1}{2} - \frac{\delta}{\sqrt{n}\gamma(x)} > 0$ then (26) holds for every $0 < \mu \leq \frac{\xi(x)2\sqrt{n}\gamma(x)}{\sqrt{n}\gamma(x)-2\delta} := \bar{\mu}(x, \delta)$. We proved (26) which implies (25).

L3. If $x_i = [M(\omega)x]_i + q_i(\omega)$, i.e. $i \in \beta(x)$ then

$$\begin{aligned}
\|[\tilde{\Phi}'_\omega(x,\mu)]_i - [V]_i\| &= \|y_i[M(\omega)]_i + (1-y_i)e^i - \lambda[M(\omega)]_i - (1-\lambda)e^i\| \\
&= \|\frac{1}{2}[M(\omega)]_i + \frac{1}{2}e^i - \lambda[M(\omega)]_i - e^i + \lambda e^i\| \\
&= \|(\frac{1}{2} - \lambda)([M(\omega)]_i - e^i)\| \\
&= |\frac{1}{2} - \lambda|\|[M(\omega)]_i - e^i\| \\
&= 0 \leq \frac{\delta}{\sqrt{n}},
\end{aligned}$$

for $\lambda = 1/2$, which implies (25).

Putting together (19)-(23) and (25) we therefore obtain

$$dist_F\left(\tilde{\Phi}'_\omega(x,\mu), \partial_C\Phi_\omega(x)\right) \leq \sqrt{\sum_{i=1}^{n}\frac{\delta^2}{n}} = \delta$$

for all $0 < \mu \leq \bar{\mu}(x, \delta)$. $\square$

We also note that since $\|A\| \leq \|A\|_F$ for an arbitrary matrix $A \in R^{n,n}$, there follows from the previous lemma that

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

for all μ such that $0 < \mu \leq \bar{\mu}(x, \delta)$.

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 using that we can determine the threshold value $\bar{\bar{\mu}}(x, \delta_1) = \min_{j=1}^{N_{\max}} \bar{\bar{\mu}}_j(x, \delta_1)$, where

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

It is easy to prove that for given $\delta_1 > 0$ and this threshold value $\bar{\bar{\mu}}(x, \delta_1)$ we have

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

for all $0 < \mu \leq \bar{\bar{\mu}}(x, \delta_1)$.

The existence of threshold value and its explicit form imply that step S2 in Algorithm 3 is well-defined. As we have mentioned before, the BB direction d^k which is used in our algorithm is feasible descent direction for the function $\hat{f}_{N_k}(x^k, \mu_k)$, so the line search in Step S5 of Algorithm 1 is also well-defined. These facts pointed out that our algorithm is well-defined. Nonmonotone line search with $\varepsilon_k > 0$ gives additional possibilities for the choice of step-length ν_k.

Lemma 4 *Let $C \subset R^n$ be a compact set. Then for every $x \in C$, $N \in \{1, 2, ..., N_{\max}\}$ and $\mu, \mu_1, \mu_2 \in R_{++}$, $\mu_1 \geq \mu_2$ 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

$$\|\Phi_{\omega^j}(x)\| \leq M_3(\omega^j), \tag{27}$$

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

$$\|\tilde{\Phi}_{\omega^j}(x, \mu)\| \leq \|\tilde{\Phi}_{\omega^j}(x, \mu) - \Phi_{\omega^j}(x)\| + \|\Phi_{\omega^j}(x)\| \leq \tilde{\kappa}\mu + M_3(\omega^j) \tag{28}$$

and (27) and (28) imply

$$\begin{aligned} |\tilde{\Theta}_{\omega^j}(x, \mu) - \Theta_{\omega^j}(x)| &= \frac{1}{2}|\|\tilde{\Phi}_{\omega^j}(x, \mu)\|^2 - \|\Phi_{\omega^j}(x)\|^2| \\ &\leq \frac{1}{2}\|\tilde{\Phi}_{\omega^j}(x, \mu) - \Phi_{\omega^j}(x)\|(\|\tilde{\Phi}_{\omega^j}(x, \mu)\| + \|\Phi_{\omega^j}(x)\|) \\ &\leq \frac{1}{2}\tilde{\kappa}\mu(\tilde{\kappa}\mu + 2M_3(\omega^j)), \end{aligned}$$

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

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

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

The next theorem states that after a finite number of iterations, the sample size $N_{\max}$ is reached and kept until the end of algorithm. It is proved in a similar way as in Krejić, Krklec [11], [12].

Theorem 1 *Suppose that assumptions A1 and A2 are true and the sequence $\{x^k\}$ generated by Algorithm 1 is bounded. Furthermore, suppose that there exist a positive constant κ and a number $n_0 \in N$ 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 for every $k \geq q$ the sample size is maximal, i.e. $N_k = N_{\max}$.*

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

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 (14), 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 $\varepsilon_\delta^{N^1}(x^k,\mu) \geq \kappa > 0$ for every $k \geq n_0$ which implies that $\tilde{\nu}_1 d\ \varepsilon_\delta^{N^1}(x^k,\mu)$ is bounded from below for all k sufficiently large. Therefore, there exists at least one p such that $dm_p < \tilde{\nu}_1 d\ \varepsilon_\delta^{N^1}(x^p,\mu)$. However, this furthermore implies that $N_{p+1} \geq N_p^+ = N_{\max}$ which is in contradiction with the current assumption that the sample size stays at N^1. Therefore, the remaining two possible scenarios are as follows:

L1 There exists $\tilde{n}$ such that $N_k = N_{\max}$ for every $k \geq \tilde{n}$.

L2 The sequence of sample sizes oscillates.

Let us suppose that scenario L2 is the one that happens. Notice that the existence of $\bar{j} \in N$ such that $N_{\bar{j}}^{\min} = N_{\max}$ would imply scenario L1. Therefore, we have that $N_k^{\min} < N_{\max}$ for every $k \in N$. 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

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

Define $S := \kappa/N_{\max}$. We know that S is a positive constant. Define also a subsequence of iterations $\{x^{s_j}\}_{j\in N} := \{x^k\}_{k\in\bar{K}}$. 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 of smoothing parameters $\{\mu_k\}_{k\in N}$ 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 $j \in N$

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

Now, we can use (14) once more and apply it to the previous inequality to obtain that for every j

$$M_0 \leq \hat{f}_{\bar{N}}(x^{s_0}, \mu_{s_0}) + \bar{\kappa}\mu_{s_0} - jS.$$

Letting $j \to \infty$, we obtain the contradiction and therefore we conclude that the only possible scenario is in fact L1, i.e. there exists iteration $\tilde{n}$ such that $N_k = N_{\max}$ for every $k \geq \tilde{n}$. $\square$

Our Algorithm 1 is constructed in a such way that it never stops. The cases 2) and 3) in Step S3 of algorithm can happen only a finite number of times, so the construction of algorithm and Theorem 1 imply that there exists $q \in N$ such that for every $k \geq q$ there follows $N_k = N_{\max}$. This implies that eventually in k-th iteration, $k \geq q$, we are solving the optimization problem with the objective function $\hat{f}_{N_{\max}}(x, \mu_k)$. So, Algorithm 1 becomes simpler and from now on we consider the Algorithm 1 with $N_k = N_{\max}$, which will be named Algorithm.

Let

$$\hat{f}_{N_{\max}}(x) = \frac{1}{N_{\max}} \sum_{j=1}^{N_{\max}} \|\Phi_{\omega^j}(x)\|^2, \tag{29}$$

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

for $x \in R^n$, $\mu \in R_{++}$ and x_q be the iteration for which the sample size $N_{\max}$ is reached and kept until the end, i.e. x_q is the iteration such that for every $k \geq q$ holds $N_k = N_{\max}$.

Algorithm

S1: If $\| \min\{x^k, \nabla \hat{f}_{N_{\max}}(x^k, \mu_k)\}\| = 0$ then $x^{k+1} = x^k$, $\mu_{k+1} = \frac{\mu_k}{2}$, $\alpha_{k+1} = \alpha_{\min}$, and go to step S7.

S2: Compute BB direction d^k by (12) and (13) using $\nabla \hat{f}_{N_{\max}}(x^k, \mu_k)$.

S3: Find the smallest nonnegative integer j such that $\nu_k = \beta^j$ satisfies

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

S4: Set $x^{k+1} = x^k + \nu_k d^k$.

S5: If

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

then

$$\tilde{\beta} = \hat{f}_{N_{\max}}(x^{k+1})$$

and choose μ_{k+1} such that

$$0 < \mu_{k+1} \leq \min\{\frac{\mu_k}{2}, \frac{\alpha\tilde{\beta}}{2\bar{\kappa}}, \bar{\bar{\mu}}(x^{k+1}, \gamma\tilde{\beta})\}$$

else

$$\mu_{k+1} = \mu_k.$$

S6: Set

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

S7: Compute α_{k+1} by (13).

S8: Set $k := k + 1$ and return to step S1.

Let us define sets

$$K = \{0\} \cup \Big\{ k, k \in N;\ k \geq q+1,\ \hat{f}_{N_{\max}}(x^k) \leq \max\{\bar{\xi}\tilde{\beta}, \frac{1}{\alpha}|\hat{f}_{N_{\max}}(x^k, \mu_{k-1}) - \hat{f}_{N_{\max}}(x^k)|\} \Big\},$$

$$K_1 = \{k, k \in K;\ \bar{\xi}\tilde{\beta} \geq \frac{1}{\alpha}|\hat{f}_{N_{\max}}(x^k, \mu_{k-1}) - \hat{f}_{N_{\max}}(x^k)|\},$$

$$K_2 = \{k, k \in K;\ \bar{\xi}\tilde{\beta} < \frac{1}{\alpha}|\hat{f}_{N_{\max}}(x^k, \mu_{k-1}) - \hat{f}_{N_{\max}}(x^k)|\}.$$

It is clear that $K = \{0\} \bigcup K_1 \bigcup K_2$.

Lemma 5 *Let $\{x^k\}$ be a sequence generated by Algorithm 1. Then the following statements hold*

a) $|\hat{f}_{N_{\max}}(x^k) - \hat{f}_{N_{\max}}(x^k, \mu_k)| \leq \alpha \hat{f}_{N_{\max}}(x^k), \quad for \quad k \geq q+1,$

b) $dist\left(\nabla \hat{f}_{N_{\max}}(x^k, \mu_k), \partial \hat{f}_{N_{\max}}(x^k)\right) \leq \gamma \hat{f}_{N_{\max}}(x^k),\ for\ k \geq q+1,\ k \in K.$

Proof. a) We can distinguish 2 cases.
Case 1. If $k \in K$ then we obtain from Lemma 4

$$|\hat{f}_{N_{\max}}(x^k) - \hat{f}_{N_{\max}}(x^k, \mu_k)| \leq \bar{\kappa}\mu_k \leq \frac{\alpha}{2}\tilde{\beta} \leq \alpha \hat{f}_{N_{\max}}(x^k).$$

Case 2. If $k \notin K$ then $\mu_k = \mu_{k-1}$, so

$$|\hat{f}_{N_{\max}}(x^k) - \hat{f}_{N_{\max}}(x^k, \mu_k)| = |\hat{f}_{N_{\max}}(x^k) - \hat{f}_{N_{\max}}(x^k, \mu_{k-1})| < \alpha \hat{f}_{N_{\max}}(x^k).$$

b) This statement follows immediately from the updating rule of smoothing parameter. □

The next theorem can be proved in the same way as Theorem 1 in Krejić, Rapajić [13].

Theorem 2 *Suppose that the assumptions of Theorem 1 are satisfied. Then there exists $q \in N$ such that the sequence $\{x^k\}_{k \geq q}$ belongs to the level set*

$$\mathcal{L}_0 = \{x \in R^n_+ :\ \ \hat{f}_{N_{\max}}(x) \leq \hat{f}_{N_{\max}}(x^q, \mu_q) + \bar{\kappa}\mu_q + \varepsilon\}. \tag{30}$$

Proof. Theorem 1 implies the existence of q such that $N_k = N_{max}$ for every $k \geq q$. Since d^k is a descent search direction, the line search implies that

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

for every $k \geq q$. Furthermore, the sequence of smoothing parameters is nonincreasing and therefore Lemma 4 implies that for every j

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

Using the previous two inequalities and the induction argument we obtain that for every j

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

Again, by using Lemma 4 and summability of the sequence $\{\varepsilon_K\}_{k \in N}$ we obtain that for any nonnegative integer j

$$\begin{aligned} \hat{f}_{N_{max}}(x^{q+j}) &\leq \hat{f}_{N_{max}}(x^{q+j}, \mu_{q+j}) + \bar{\kappa}\mu_{q+j} \\ &\leq \hat{f}_{N_{max}}(x^q, \mu_q) + \bar{\kappa}(\mu_q - \mu_{q+j}) + \varepsilon + \bar{\kappa}\mu_{q+j} \\ &= \hat{f}_{N_{max}}(x^q, \mu_q) + \bar{\kappa}\mu_q + \varepsilon. \end{aligned}$$

This completes the proof. □

Theorem 1 states that after a finite number of iterations, the sample size $N_{\max}$ is reached and kept until the end. So, our problem becomes of the form

$$\hat{f}_{N_{\max}}(x) \to \min, \quad x \geq 0, \tag{31}$$

where $\hat{f}_{N_{\max}}(x)$ is a nonsmooth function defined by (29). We want to show that every accumulation point of the sequence generated by our algorithm is a Clarke stationary point of $\hat{f}_{N_{\max}}(x)$.

As we have mentioned before, instead of solving this nonsmooth problem (31) we are solving the sequence of smooth problems

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

Using a similar idea as in Krejić, Rapajić [13] the following theorem can be proved.

Theorem 3 *Let $\{x^k\}$ be a sequence generated by Algorithm 1. Suppose that assumptions A1-A2 are satisfied and the level set (30) is bounded. Then every accumulation point of the sequence $\{x^k\}$ is a Clarke stationary point of $\hat{f}_{N_{\max}}(x)$.*

Proof. Let x^* be an accumulation point of $\{x^k\}$ and $\{x^k\}_L$, $L \subseteq N_1 = \{k,\ k \in N,\ k \geq q+1\}$ be a subsequence which converges to x^*, i.e.

$$\lim_{k\to\infty, k\in L} x^k = x^*. \tag{32}$$

For x^* let us denote

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

From Definitions 2-4 and the fact that $G_{\hat{f}_{N_{\max}}}(x^*) \subseteq \partial \hat{f}_{N_{\max}}(x^*)$ it is clear that if x^* is a stationary point of function $\hat{f}_{N_{\max}}(x)$ associated with a smoothing function $\hat{f}_{N_{\max}}(x,\mu)$, then x^* is a Clarke stationary point of function $\hat{f}_{N_{\max}}(x)$. Therefore, it is sufficient to prove that x^* is a stationary point of function $\hat{f}_{N_{\max}}(x)$ associated with a smoothing function $\hat{f}_{N_{\max}}(x,\mu_k)$, $\mu_k \to 0$.

If $\| \min\{x^k, \nabla \hat{f}_{N_{\max}}(x^k,\mu_k)\}\| = 0$, for all $k \in L$ then by (11)

$$\langle \nabla \hat{f}_{N_{\max}}(x^k,\mu_k), x - x^k\rangle \geq 0, \tag{33}$$

for any $x \geq 0$. By Definition 1, there exists an infinite subset $L_1 \subseteq L$ such that $\lim_{k\to\infty, k\in L_1} \nabla \hat{f}_{N_{\max}}(x^k,\mu_k) = V \in G_{\hat{f}_{N_{\max}}}(x^*)$, because $\mu_k \to 0$. From (32) and (33) there follows

$$\lim_{k\to\infty, k\in L_1} \langle \nabla \hat{f}_{N_{\max}}(x^k,\mu_k), x - x^k\rangle = \langle V, x - x^*\rangle \geq 0,$$

for any $x \geq 0$ which by Definition 4 means that x^* is a stationary point of function $\hat{f}_{N_{\max}}(x)$ associated with a smoothing function $\hat{f}_{N_{\max}}(x,\mu_k)$, thus implies that x^* is a Clarke stationary point of $\hat{f}_{N_{\max}}(x)$.

Otherwise, let us assume that there exists $L_1 \subseteq L$ such that for some $\tilde{\varepsilon} > 0$ we have

$$\| \min\{x^k, \nabla \hat{f}_{N_{\max}}(x^k,\mu_k)\}\| \geq \tilde{\varepsilon} > 0, \quad k \in L_1. \tag{34}$$

The following two cases are considered separately:

Case 1: K is finite set.

Case 2: K is infinite set.

Case 1. Let $\hat{k}$ be the largest index in K. Then $\mu_k = \mu_{\hat{k}}$, for every $k \geq \hat{k}$. Without loss of generality, it can be assumed that $k \notin K$, for all $k \in L_1$, because K is finite set. The level set is bounded, so $\nabla \hat{f}_{N_{\max}}(x^k, \mu_k)$ is also bounded and thus the direction d^k is bounded. As $\lim_{k\to\infty, k\in L_1} x^k = x^*$ and $\alpha_k \geq \alpha_{\min}$, we have $\lim_{k\in L_1, k\to\infty} \nu_k d^k = 0$. Therefore, we consider two different possibilities:

a) $\nu_k \to 0$, $k \in L_2$, for some $L_2 \subseteq L_1$,

b) $\nu_k \geq \nu_* > 0$, $k \in L_2$.

a) If $\nu_k \to 0$, $k \in L_2$, the choice of step-length in the line-search implies that for each $k \in L_2$ there exists $\nu'_k > \nu_k$, such that

$$\lim_{k\in L_2} \nu'_k = 0$$

and

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

so

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

Therefore, taking the limits we obtain

$$\lim_{\nu'_k \to 0} \frac{\hat{f}_{N_{\max}}(x^k + \nu'_k d^k, \mu_k) - \hat{f}_{N_{\max}}(x^k, \mu_k)}{\nu'_k} \geq \eta (d^k)^T \nabla \hat{f}_{N_{\max}}(x^k, \mu_k)$$

i.e.

$$(d^k)^T \nabla \hat{f}_{N_{\max}}(x^k, \mu_k) \geq \eta (d^k)^T \nabla \hat{f}_{N_{\max}}(x^k, \mu_k).$$

Given that d^k is a descent direction , the last inequality implies $\eta \geq 1$, which is not possible and thus K can not be a finite set.

b) If $\nu_k \geq \nu_* > 0$ for $k \in L_2$, then as $k \in L_2$ we may assume that $k \notin K$ and $\mu_k = \mu_{\hat{k}} = \mu$. Then

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

$$
\begin{aligned}
&\leq \hat{f}_{N_{\max}}(x^{k-1},\mu)+\eta\nu_{k-1}(d^{k-1})^T\nabla\hat{f}_{N_{\max}}(x^{k-1},\mu)+\varepsilon_{k-1}\\
&\quad +\eta\nu_k(d^k)^T\nabla\hat{f}_{N_{\max}}(x^k,\mu)+\varepsilon_k\\
&\leq \cdots\\
&\leq \hat{f}_{N_{\max}}(x^{\hat{k}+1},\mu)+\eta\sum_{j=\hat{k}+1}^{k}\nu_j(d^j)^T\nabla\hat{f}_{N_{\max}}(x^j,\mu)+\sum_{j=\hat{k}+1}^{k}\varepsilon_j.
\end{aligned}
$$

So, using $\nu_k \geq \nu_* > 0$ for $k \in L_2$ we get

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

Given that $\hat{f}_{N_{\max}}(x,\mu)$ is bounded for $x \in \mathcal{L}_0$ and $\mu > 0$ and $\sum_{k\in N}\varepsilon_k \leq \varepsilon$, there follows that

$$
\lim_{k\to\infty,k\in L_2}(d^k)^T\nabla\hat{f}_{N_{\max}}(x^k,\mu)=0. \tag{35}
$$

On the other hand, the definition of d^k implies

$$
d_i^k[\nabla\hat{f}_{N_{\max}}(x^k,\mu)]_i=\begin{cases}
-\frac{1}{\alpha_k}[\nabla\hat{f}_{N_{\max}}(x^k,\mu)]_i^2, & \text{if } i\in I_1(x^k)\\
-\frac{[\nabla\hat{f}_{N_{\max}}(x^k,\mu)]_i^2}{\alpha_k+\frac{[\nabla\hat{f}_{N_{\max}}(x^k,\mu)]_i}{x_i^k}}, & \text{if } i\in I_2(x^k)\\
-x_i^k[\nabla\hat{f}_{N_{\max}}(x^k,\mu)]_i, & \text{if } i\in I_3(x^k)
\end{cases},
$$

so

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

Given that α_k is bounded and that each of the three sums above is non-positive, (35) implies

$$
\lim_{k\to\infty,k\in L_2}[\nabla\hat{f}_{N_{\max}}(x^k,\mu)]_i=0,\quad i\in I_1(x^k)\cup I_2(x^k)
$$

and

$$
\lim_{k\to\infty,k\in L_2}x_i^k[\nabla\hat{f}_{N_{\max}}(x^k,\mu)]_i=0,\quad i\in I_3(x^k).
$$

Thus

$$\lim_{k\to\infty,k\in L_2} \min\{x_i^k, [\nabla \hat{f}_{N_{\max}}(x^k,\mu)]_i\} = 0$$

and

$$\lim_{k\to\infty,k\in L_2} \| \min\{x^k, \nabla \hat{f}_{N_{\max}}(x^k,\mu)\}\| = 0,$$

which is a contradiction with (34) and means that K can not be a finite set.

Case 2. The set K is infinite, so from the updating rule of smoothing parameter, there follows that

$$\lim_{k\to\infty,k\in L_1} \mu_k = 0. \tag{36}$$

As $\lim_{k\to\infty,k\in L_1} x^k = x^*$ and K is infinite set, without loss of generality it can be assumed that $k \in K$, for every $k \in L_1$ large enough. The set K is infinite, so $k_j \to \infty$ and $j \to \infty$. Then from (**??**) there follows

$$0 \le \hat{f}_{N_{\max}}(x^k) \le r^j(1+\alpha)\hat{f}_{N_{\max}}(x^0) + \varepsilon_{k_j},$$

for $j \ge 1$. Therefore, taking limits we have

$$0 \le \hat{f}_{N_{\max}}(x^*) = \lim_{k\to\infty,k\in L_1} \hat{f}_{N_{\max}}(x^k) \le \lim_{j\to\infty} (r^j(1+\alpha)\hat{f}_{N_{\max}}(x^0) + \varepsilon_{k_j}) = 0,$$

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 will 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 way of updating the smoothing parameter is more complex than in LLS.

In order to provide a better insight into the results, we will state the relevant notation considering the test problems. The point $\hat{x}$ is defined in a way to poses exactly n_x positive components which we choose 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 represents uniformly distributed random vector. The stopping criterion for the both algorithms is

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

with $\gamma = 100$ and $\epsilon = 10^{-6}$. The search direction d^k is determined by using $\alpha_{min} = \alpha_0 = 0.1$. Line search is performed with both β and η equal to 0.5. The starting smoothing parameter is $\mu_0 = 1$. In LLS, this parameter is updated by multiplying with 0.5, 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 the relevant algorithms. The results presented in the following two tables represent average values of successful runs reported in column s. The 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 $\nabla_x F$. More precisely, each component of the gradient is counted as one function evaluation. The column *stac* refers to the measure of stationarity $\| \min\{x^k, \nabla \hat{f}_{N_{max}}(x^k, \mu_k)\} \|$.

$\beta_e = 0$	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 represents 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. Instances 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 we can not say which one of the tested algorithms performs better. However, the results in Table 2 reveal the clear advantage of VSS method. In all the tested problems presented in the last table, average number of function evaluations for the VSS is lower than fev for LLS. Therefore, our conclusion is that the overall results suggest that using the algorithm VSS can be beneficial.

References

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

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

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

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

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

[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. pp. 1-32, 2009.

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

[10] C. Kanzow and H. Pieper, Jacobian smoothing methods for general nonlinear complementarity problems, SIAM Journal on Optimization 9, pp. 342-373, 1999.

[11] N. Krejić and N. Krklec, Line search methods with variable sample size for unconstrained optimization, Journal of Computational and Applied Mathematics 245, pp. 213-231, 2013.

[12] N. Krejić and N. Krklec, Nonmonotone line search methods with variable sample size, technical report, 2013.

[13] N. Krejić and S. Rapajić, Globally convergent Jacobian smoothing inexact Newton methods for NCP, Computational Optimization and Applications 41, pp. 243-261, 2008.

[14] C. Zhang and X. Chen, Smoothing projected gradient method and its application to stochastic linear complementarity problems, Siam. J. Optim. Vol. 20, No. 2, pp. 627-649, 2009.