Numerical Algorithms manuscript No.
(will be inserted by the editor)

Penalty variable sample size method for solving optimization problems with equality constraints in a form of mathematical expectation

Nataša Krklec Jerinkić · **Andrea Rožnjik**

Received: date / Accepted: date

Abstract Equality constrained optimization problems with deterministic objective function and constraints in the form of mathematical expectation are considered. The constraints are approximated by employing the sample average where the sample size varies throughout the iterations in an adaptive manner. The proposed method incorporates variable sample size scheme with cumulative and unbounded sample into the well known quadratic penalty iterative procedure. Line search is used for globalization and the sample size is updated in a such way to preserve the balance between two types of errors - errors coming from the sample average approximation and the approximation of the optimal point. Moreover, the penalty parameter is also updated in an adaptive way. We prove that the proposed algorithm pushes the sample size and the penalty parameter to infinity which further allows us to prove the almost sure convergence towards a Karush-Kuhn-Tucker optimal point of the original problem under the rather standard assumptions. Numerical comparison on a set of relevant problems shows the advantage of the proposed adaptive scheme over the heuristic (predetermined) sample scheduling in terms of number of function evaluations as a measure of the optimization cost.

Keywords stochastic optimization · equality constraints · sample average approximation · variable sample size · quadratic penalty method · line search

N. Krklec Jerinkić is supported by Serbian Ministry of Education, Science and Technological Development, grant no. 174030.

Nataša Krklec Jerinkić
Department of Mathematics and Informatics, University of Novi Sad,
Trg Dositeja Obradovića 4, 21000 Novi Sad, Serbia
E-mail: natasa.krklec@dmi.uns.ac.rs

Andrea Rožnjik
Faculty of Civil Engineering, University of Novi Sad,
Kozaračka 2a, 24000 Subotica, Serbia
E-mail: andrea@gf.uns.ac.rs

1 Introduction

We consider equality constrained problems

$$\min_x f(x) \quad \text{subject to } h(x) = 0, \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is bounded from below and the constraints are assumed to be in the form of mathematical expectation. More precisely,

$$h(x) = E[H(x, \xi)],$$

where ξ represents a random vector defined on a probability space $(\Omega, \mathscr{F}, P)$ and $H : \mathbb{R}^n \times \Omega \to \mathbb{R}^m$. We state the needed assumptions regarding the objective function and the constraints below.

Assumption 1 *Function f is continuously differentiable and bounded from below. Moreover, $H(\cdot, \xi) \in C^1(\mathbb{R}^n)$ for every ξ.*[1]

Stochastic optimization problems arise from various scientific fields and they have been studied extensively in a past few decades (see [4,22,23] for further references). The stochastic nature of the relevant functions is often removed by employing mathematical expectation and the stochastic problem is then transformed into the deterministic one. However, analytical form of the mathematical expectation is rarely available and it is often approximated by the sample average yielding the so called Sample Average Approximation (SAA) methods [9,10,15,19–22,24]. Since the sample needed for a reasonable approximation of the original function is usually very large, evaluations of the sample average functions tend to be costly. This motivated the development of Variable Sample Size (VSS) methods such as [1–3,5,8, 11–14]. The main reasoning behind the VSS approach is to use smaller samples in the early iterations, while the vicinity of the solution is still to be reached, and save some costs by saving (usually significant) number of function evaluations. Although heuristic approaches such as increasing the sample size at every iteration are straight forward and easy to implement, recent numerical studies indicate that an adaptive sample size scheduling may improve the performance of the algorithm significantly [11–14]. The method proposed in this paper uses an adaptive VSS scheme similar to one developed in [11,12] for line search methods (its trust region counterparts are developed earlier in Bastin et al. [1–3]).

Methods presented in [11,12] are developed to solve only the SAA variant of an unconstrained problem where the objective function is an SAA estimator of the mathematical expectation. Namely, although the motivation comes from the problems where the objective function is in a form of mathematical expectation, it is assumed therein that the (large, but finite) sample is generated in advance. This practically yields deterministic optimization problem and the algorithms are constructed in a way to reach the full sample eventually and provide classical (deterministic) convergence results. The method proposed in this paper differs in several ways. The problem that

[1] In order to obtain the final convergence result stated in Theorem 3, one may relax this assumption and state "almost every" instead of "every" ξ (see the remark after the proof of Theorem 3).

we observe is a constrained optimization problem. Although the objective function is not explicitly in a form of mathematical expectation, the analysis can be extended to this case as well. However, we decided to focus on the constraints as in [14] where the SAA variant of problem (1) is solved.

In this paper, we aim to solve the original problem (1) by using a sequence of SAA estimators. This leads us to an unbounded sample and stochastic convergence theoretical results and requests nontrivial modifications (with respect to [14]) concerning both convergence analysis and the construction of algorithm itself. Similar transfer from bounded to unbounded sample case was done in [13]. However, in that paper, constraints are assumed to be deterministic and easy to project on which is not the case here. Instead, we use the quadratic penalty method to cope with the constraints as it was done in [14]. We use the SAA estimator of the constraints $h(x)$ given by

$$h_N(x) = \frac{1}{N}\sum_{i=1}^{N} H(x,\xi_i), \tag{2}$$

where $\xi_1, \xi_2, \ldots, \xi_N$ represent a sample realization with $N \in \mathbb{N}$ being the sample size. At each iteration we use potentially different sample size, but we assume that the sample is cumulative. Therefore, although the stochastic nature of the algorithm yields different sample paths, the sample size uniquely determines the sample within one sample path and thus we may use the SAA form defined above.

We use the quadratic penalty function

$$\phi_N(x;\mu) = f(x) + \mu\,\theta_N(x)$$

where the (approximate) measure of infeasibility is given by [2]

$$\theta_N(x) := \|h_N(x)\|^2$$

and μ represents the penalty parameter. At each iteration we update: a) the decision variable x; b) the sample size N; c) the penalty parameter μ. As in the classical penalty method, μ needs to be pushed to infinity. This is ensured by an adaptive rule similar to one in [14]. Furthermore, in order to reach the mathematical expectation at limit, we need to push N to infinity as well. This is also done adaptively and this result makes one of the key points of our analysis. Having the sample size and the penalty parameter tend to infinity, we prove that a Karush-Kuhn-Tucker (KKT) point of the problem (1) is attainable under rather standard assumptions.

At every iteration k, line search backtracking technique with respect to the decision variable x is applied on the penalty function $\phi_{N_k}(x_k;\mu_k)$ along the search direction d_k which is assumed to be descent. More precisely, we assume that d_k satisfies the following two inequalities

$$d_k^T g_k \leq -\lambda_{min}\|g_k\|^2, \tag{3}$$

$$\|d_k\| \leq \lambda_{max}\|g_k\| \tag{4}$$

[2] We use $\|\cdot\|$ to denote Euclidean norm.

where λ_{min} and λ_{max} are some positive constants and

$$g_k := \nabla_x \phi_{N_k}(x_k; \mu_k).$$

One possible approach is to set $d_k = -B_k g_k$ where B_k is a symmetric matrix, uniformly bounded and uniformly positive definite across all iterations. Obviously, the steepest descent direction satisfies (3)-(4). Moreover, the spectral gradient method with the projection of the spectral coefficient on the interval $[\lambda_{min}, \lambda_{max}]$ also makes a feasible choice. More generally, a Quasi-Newton method with a suitable safeguarding would also be applicable (see [16] for further references). For example, one may employ some update-type (inverse) Hessian approximation method such as BFGS and skip the update if the eigenvalues of the proposed matrix B_{k+1} fall out of the interval $[\lambda_{min}, \lambda_{max}]$. Notice that (3) implies that g_k tends to zero if $d_k^T g_k$ does. Moreover, (4) implies that the sequence of d_k is uniformly bounded if $\{g_k\}$ is.

Since the search direction is assumed to be descent, the monotone line search is applicable. Although nonmonotone line search methods could be employed, we focus on Armijo line search for simplicity. More precisely, given the search direction d_k, the step size α_k is determined such that

$$\phi_{N_k}(x_k + \alpha_k d_k; \mu_k) \leq \phi_{N_k}(x_k; \mu_k) + \eta \alpha_k g_k^T d_k, \tag{5}$$

fore some $\eta \in (0,1)$. This means that the penalty function is decreased for some portion of the scalar product on the right hand side of the previously stated inequality. We denote the measure of decrease by

$$dm_k := -\alpha_k g_k^T d_k. \tag{6}$$

Notice that $dm_k \geq 0$. This measure plays an important role in the sample size update. It can be viewed as a measure of distance of x_k from the stationary point of $\phi_{N_k}(\cdot; \mu_k)$. On the other hand, SAA estimator (2) produces another type of error. We denote the measure of this error by $e(x;N)$. The main idea behind the sample size update is to keep these two types of errors balanced. If the stationary point of the current penalty function is approached (that is, if dm_k is relatively small), we increase the precision of the SAA by increasing the sample size. This mechanism is similar to one in the so-called diagonalization, [17,18], where the sequence of SAA problems is solved. However, the method that we use differs in a way that it allows the decrease of the sample size if the stationary point is still far away in order to diminish the optimization costs. The sample size update is similar to one in [14], but it is stated for completeness in Algorithm 2 (Step 1). The main difference lies in Step 2 where the sample size lower bound N_k^{min} is updated. This part is modified to cope with unbounded samples and it plays an important role in the convergence analysis.

The remainder of this paper is organized as follows. Details of the proposed algorithm are presented in the next section. Section 3 provides convergence analysis and numerical results are presented in Section 4. Some conclusions are drawn in the final section.

2 The Algorithm

We start this section by stating the framework algorithm while the sample size update is stated in Algorithm 2.

Algorithm 1
Step 0 Input parameters: $N_0 \in \mathbb{N}$, $x_0 \in \mathbb{R}^n$, $\beta, \eta \in (0,1)$, $\mu_0 > 0$, $\rho > 1$.
Step 1 Set $k = 0$, $N_k = N_0$, $x_k = x_0$, $\mu_k = \mu_0$, $t = 1$, $N_0^{min} = N_0$.
Step 2 If

$$g_k = 0 \quad \text{and} \quad h_{N_k}(x_k) = 0 \tag{7}$$

set

$$N_{k+1} = N_k + 1 \ \text{ and } \ N_{k+1}^{min} = \max\{N_{k+1}, N_k^{min} + 1\}$$

and go to Step 7.
Step 3 Determine the search direction d_k that satisfies (3)-(4).
Step 4 Determine the step size α_k:
Find the smallest nonnegative integer j such that $\alpha_k = \beta^j$ satisfies (5).
Set $x_{k+1} = x_k + \alpha_k d_k$ and $dm_k = -\alpha_k g_k^T d_k$.
Step 5 Determine the sample size N_{k+1} by applying Algorithm 2.
Step 6 Determine the penalty parameter μ_{k+1}:
If

$$dm_k \leq \frac{\alpha_k}{\mu_k^2},$$

set $\mu_{k+1} = \rho\mu_k$, $z_t = x_k$ and $t = t + 1$.
Else set $\mu_{k+1} = \mu_k$.
Step 7 Set $k = k + 1$ and go to Step 2.

Assumption 1 ensures that the Algorithm 1 is well defined. It implies that h_N is continuously differentiable for every $N \in \mathbb{N}$ and thus the penalty function is also continuously differentiable, i.e. g_k is available at every iteration.

Notice that, besides in Step 5, the sample size can be updated in Step 2 of Algorithm 1 as well. The reasoning behind this additional update lies in the fact that (7) implies that x_k is a KKT point of the relevant SAA problem

$$\min_x f(x) \quad \text{subject to } \ h_{N_k}(x) = 0,$$

and thus the higher level of precession is to be used in the next step. Indeed, since

$$g_k = \nabla f(x_k) + 2\mu_k \nabla^T h_{N_k}(x_k) h_{N_k}(x_k),$$

the KKT conditions are satisfied with the Lagrange multiplier $\lambda_k = 2\mu_k h_{N_k}(x_k)$.

In Step 3 we calculate the descent search direction and in Step 4 we determine the step size via Armijo line search. The objective function is assumed to be bounded from below and the penalty part is nonnegative, so the penalty function is also bounded from below and the line search is well defined. Moreover, since the search direction satisfies (4), $g_k = 0$ implies $d_k = 0$ and the Armijo condition is trivially satisfied with the full step. Within this step we also calculate dm_k needed for determining the subsequent sample size.

The main sample size update is performed in Algorithm 2 stated below where the lower bound sample size N_{k+1}^{min} is also determined. Finally, in Step 6 we update the penalty parameter. If the measure of stationarity related to the current penalty function is small enough, we increase the penalty parameter and update the sequence $\{z_t\}$. This sequence is actually a subsequence of $\{x_k\}$ and it plays an important role in convergence analysis. More precisely, we prove that, under rather standard assumptions, the sequence $\{z_t\}$ is infinite and it converges towards a KKT point of the original problem almost surely.

Next, we state the algorithm for the sample size update. Although it is similar to one in [14] and [11,12], we state it here for completeness and give a brief description. Assumption concerning SAA error measure $e(x,N)$ is given below.

Assumption 2 *Function $e : \mathbb{R}^n \times \mathbb{N} \to \mathbb{R}_+$ is such that for any finite valued N there exists e_N such that $e(x;N) \geq e_N > 0$ for every $x \in \mathbb{R}^n$.*

Assumption 2 is as in [13]. It is rather mild and it allows many different choices for the function e, including $e(x;N) = 1/N$. Moreover, since we also update the sample size lower bound in Algorithm 2, we provide the assumption on the relevant mapping $\gamma(N)$ presented in [13] as well.

Assumption 3 $\gamma : \mathbb{N} \to (0,1)$ *is an increasing function such that* $\lim_{N\to\infty} \gamma(N) = 1$.

Notice that $\gamma(N) = \exp(-1/N)$ is one suitable choice.

Algorithm 2

Step 0 Input parameters: dm_k, $e(x_k;N_k)$, x_k, N_k, N_k^{min}.

Step 1 Determine N_{k+1}:

1) If $dm_k = e(x_k;N_k)$, set $N_{k+1} = N_k$.

2) If $dm_k > e(x_k;N_k)$, set $N = N_k$.
While $dm_k > \dfrac{N_k}{N}\, e(x_k;N)$ and $N > N_k^{min}$ set $N = N-1$.
Set $N_{k+1} = N$.

3) If $dm_k < e(x_k;N_k)$, set $N = N_k + 1$.
While $dm_k < \dfrac{N_k}{N}\, e(x_k;N)$ and $\|h_N(x_k)\| < \dfrac{N_k}{N(N-N_k)}\, e(x_k;N)$ set $N = N+1$.
Set $N_{k+1} = N$.

Step 2 Determine N_{k+1}^{min}:

If $N_{k+1} \neq N_k$ and there exists $i \in \{1,..,k-1\}$ such that $N_i = N_{k+1}$, determine

$$l(k) = \max\{i \in \{1,\ldots,k-1\} \mid N_i = N_{k+1}, N_{i-1} \neq N_{k+1}\}.$$

If

$$\frac{\theta_{N_{k+1}}(x_{l(k)}) - \theta_{N_{k+1}}(x_{k+1})}{k+1-l(k)} < \gamma(N_{k+1})\, e(x_{k+1};N_{k+1}), \tag{8}$$

set $N_{k+1}^{min} = \max\{N_{k+1}, N_k^{min}+1\}$.
Else $N_{k+1}^{min} = N_k^{min}$.

In Step 1 we update the sample size such that two error measures - dm_k and $e(x_k;N_{k+1})$ - are balanced. The first one is defined above in (6) and the second one represents the error of approximation (2) with properties defined in Assumption 2. For example, one can view $e(x_k;N_{k+1})$ as the loglog bound proposed in [8] for cumulative samples. However, numerical study implies that less conservative (for example, variance-related) bound works well in practice. Theoretically, key issues provided in this step are:

a) N_{k+1} is strictly larger than N_k if $dm_k < e(x_k;N_k)$;
b) $N_k \geq N_k^{min} - 1$ for all $k = 0, 1, \ldots$.

The first issue (a)) is provided by Step 1 3) where the sample size is increased until dm_k and $e(x_k;N_{k+1})$ are in balance or until the SAA error reaches the measure of infeasibility. Adding the second condition is motivated as follows - it aims to prohibit unproductively large increase of the sample size. The ratio N_k/N_{k+1} present in Step 1 2) and Step 1 3) is here to motivate less changes on smaller samples when the accuracy is not that good and to allow greater leaps when the sample is already large and the solution is probably approached. The ratio $N_k/(N_{k+1}(N_{k+1}-N_k))$ in Step 1 3) shares the same role.

The main reasoning behind Step 2 is to increase the lower level of precision (that is, to set $N_{k+1}^{min} = \max\{N_{k+1}, N_k^{min}+1\}$) if the decrease of infeasibility measure related to the sample size N_{k+1} (that is, to $\theta_{N_{k+1}}$) is not satisfactory. The left hand side of (8) represents average decrease from the iteration at which the same level of precision N_{k+1} was used. More precisely, $l(k)$ represents the iteration at which we started to use N_{k+1} for the last time before iteration $k+1$. Notice that $l(k)$ is calculated only if there exists $i \in \{1, \ldots, k-1\}$ such that $N_i = N_{k+1}$ as stated in Step 2. Otherwise, the lower bound is not altered. The right hand side of (8) represents a "sufficient" decrease. It is given by the product of the SAA error measure $e(x_{k+1};N_{k+1})$ and mapping γ determined by Assumption 3. The key property of this mapping is that it is increasing, positive and bounded from above. The bottom line of this step is that the obtained decrease is compared to the SAA error - similar to Step 1 - and the role of γ is just provide a more rigorous treatment of higher levels of precision in order to improve the precision even more.

Now, let us show that the second issue (b)) holds. Initially we set $N_0^{min} = N_0$, so the inequality is obviously true for $k = 0$. Let us assume that $N_k \geq N_k^{min} - 1$ and observe Algorithm 2. Considering Step 1, we have the following cases.

1) If the sample size is unchanged, then the same is true for N_k^{min} and we have

$$N_{k+1} = N_k \geq N_k^{min} - 1 = N_{k+1}^{min} - 1.$$

2) If $N_k = N_k^{min} - 1$, then the same analysis as in Step 1) applies. Otherwise, if the sample size decrease is attempted, the algorithm ensures that $N_{k+1} \geq N_k^{min}$. Moreover, according to Step 2 we obtain

$$N_{k+1}^{min} \leq \max\{N_{k+1}, N_k^{min}+1\} \leq \max\{N_{k+1}+1, N_k^{min}+1\} = N_{k+1}+1,$$

which obviously implies $N_{k+1} \geq N_{k+1}^{min} - 1$.

3) If the sample size is increased, then we know that $N_{k+1} \geq N_k + 1$ and the assumption yields $N_k + 1 \geq N_k^{min}$, so there holds $N_{k+1} \geq N_k^{min}$. Now, Step 2 yields the following possibilities.
   i) If $N_{k+1}^{min} = N_{k+1}$, then obviously $N_{k+1} \geq N_{k+1}^{min} - 1$.
   ii) If $N_{k+1}^{min} = N_k^{min} + 1$, then we obtain the same result as

$$N_{k+1} \geq N_k^{min} = N_{k+1}^{min} - 1.$$

   iii) If $N_{k+1}^{min} = N_k^{min}$, then

$$N_{k+1} \geq N_k^{min} = N_{k+1}^{min} \geq N_{k+1}^{min} - 1.$$

Finally, if the sample size is updated within Step 2 of Algorithm 1, the analysis similar to one in part 3) shows that $N_{k+1} \geq N_{k+1}^{min} - 1$. Therefore, b) holds and we conclude that the sequence $N_k^{min} - 1$ is the lower bound sequence of N_k.

At the end of this section, notice that Algorithms 1 - 2 ensure that the sequences $\{\mu_k\}$ and $\{N_k^{min}\}$ are nondecreasing.

3 Convergence analysis

We start the analysis by proving that the above algorithms push the sample size sequence to infinity. The proof of Lemma 1 follows the ideas from [14]. It represents an intermediate result which states that the sequence of sample sizes cannot become stationary. The key point lies in Step 1 of Algorithm 2. On the other hand, the result of Theorem 1 heavily leans on Step 2.

Remark 1 The sequences generated by the main algorithm are obviously random. On the other hand, the following two results are stated in deterministic manner. This is possible because they hold for any given sample path. Of course, almost every object within the proofs is random, i.e., sample path-dependent, but the lines in the proofs hold surely since they fix an arbitrary sample path and observe what happens within it.

Lemma 1 *Suppose that Assumptions 1 - 2 hold. Then there cannot exist $q \in \mathbb{N}$ such that $N_{k+1} = N_k$ for every $k \geq q$.*

Proof Let us assume a contrary, i.e., suppose that there are $\bar{q}, \bar{N} \in \mathbb{N}$ such that

$$N_{k+1} = N_k = \bar{N} \quad \text{for every} \quad k \geq \bar{q}. \tag{9}$$

We will prove that this implies that

$$\liminf_{k \to \infty} dm_k = 0. \tag{10}$$

Consider the penalty parameter. There are two possibilities:

P1 μ_k changes finitely many times,
P2 μ_k changes infinitely many times.

Having in mind that the sequence of penalty parameters is nondecreasing, P1 means that there exists $\bar{\mu}$ such that $\mu_k = \bar{\mu}$ for every k large enough. Without loss of generality, we can say that $\mu_k = \bar{\mu}$ for every $k \geq \bar{q}$. Consequently, denoting $\phi(x_k) := \phi_{\bar{N}}(x_k;\bar{\mu})$, Step 4 of Algorithm 1 implies that

$$\phi(x_{k+1}) \leq \phi(x_k) - \eta\, dm_k$$

for every $k \geq \bar{q}$. Furthermore, Assumption 1 implies the existence of a constant M such that $\phi(x_k) \geq M$ for every k and thus we obtain

$$\sum_{k=\bar{q}}^{\infty} dm_k \leq \frac{1}{\eta}\left(\phi(x_{\bar{q}}) - M\right) < \infty.$$

Therefore, there holds that $\lim_{k\to\infty} dm_k = 0$.

On the other hand, since changing μ_k actually means increasing μ_k, P2 implies that $\lim_{k\to\infty} \mu_k = \infty$. Let $K \subseteq \mathbb{N}$ be an infinite subset of iterations at which the penalty parameter is increased, i.e., $\mu_{k+1} = \rho\mu_k$ for every $k \in K$. Since the step size α_k is bounded from above, Step 6 of Algorithm 1 furthermore implies that $\lim_{k\in K} dm_k = 0$.

We have just proved that both P1 and P2 imply (10). On the other hand, since the sample size is assumed to be fixed, Assumption 2 implies that $e(x_k;N_k)$ is bounded away from zero. More precisely, $e(x_k;N_k) \geq e_{\bar{N}} > 0$ for every $k \geq \bar{q}$. Therefore, (10) implies the existence of finite iteration $s \geq \bar{q}$ such that $dm_s < e_{\bar{N}} \leq e(x_s;\bar{N})$, so Step 1 3) of Algorithm 2 implies that $N_{s+1} > N_s$ which is in contradiction with (9). This completes the proof. □

Next, we prove that the sample size sequence tends to infinity under the additional assumption on γ, i.e., Assumption 3.

Theorem 1 *Suppose that Assumptions 1-3 hold. Then*

$$\lim_{k\to\infty} N_k = \infty.$$

Proof If the sample size N_k changes in Step 2 of Algorithm 1 infinitely many times, then the statement trivially holds since N_k^{min} tends to infinity. Thus, let us assume that Step 2 activates the increase only finitely many times, i.e., there exists $q \in \mathbb{N}$ such that for every $k \geq q$ conditions (7) are not fulfilled and the sample size update can only be conducted within Algorithm 2.

Let us consider the lower bound sequence $\{N_k^{min}\}_{k\geq q}$. Recall that this sequence is nondecreasing and $N_k \geq N_k^{min} - 1$, so if N_k^{min} tends to infinity, the results holds as stated above. Therefore, let us consider the case where the sequence $\{N_k^{min}\}_{k\geq q}$ is bounded. That means that there are $\bar{q} \geq q$ and $N_{max}^{min} \in \mathbb{N}$ such that

$$N_k^{min} = N_{max}^{min} \quad \text{for every} \;\; k \geq \bar{q}. \tag{11}$$

Now, suppose that the statement of this theorem does not hold, i.e., suppose that there is a bounded infinite subsequence of $\{N_k\}_{k\in\mathbb{N}}$. This furthermore implies the existence of $\bar{N} \in \mathbb{N}$ and infinite subsequence $\{\bar{k}_1, \bar{k}_2, \ldots\} \subset \{k \in \mathbb{N} : k \geq \bar{q}\}$ such that $N_{\bar{k}_i+1} = \bar{N}$

for $i = 1,2,\ldots$. Due to Lemma 1, the sequence of sample sizes cannot be stationary and thus there must exist an infinite subsequence $\{k_1,k_2,\ldots\} \subset \{\bar{k}_1,\bar{k}_2,\ldots\}$ such that

$$\bar{N} = N_{k_i+1} \neq N_{k_i} \quad \text{for } i = 1,2,\ldots.$$

Since we assumed (11), Step 2 of Algorithm 2 implies that

$$\frac{\theta_{\bar{N}}(x_{l(k_i)}) - \theta_{\bar{N}}(x_{k_i+1})}{k_i + 1 - l(k_i)} \geq \gamma(\bar{N})\, e(x_{k_i+1};\bar{N})$$

for every $i = 1,2,\ldots$. Let us use the notation $k_i^+ := k_i + 1$ for convenience. Notice that $l(k_i) = k_{i-1}^+$ and the previous inequality together with Assumptions 2-3 implies that

$$\theta_{\bar{N}}(x_{k_{i-1}^+}) - \theta_{\bar{N}}(x_{k_i^+}) \geq \gamma(\bar{N}) e_{\bar{N}} := C > 0,$$

for every $i = 2,3,\ldots$. However, this is in contradiction with the fact that θ_N is bounded from below, so there cannot exist a bounded subsequence of sample sizes and the statement of this theorem is proved. □

In order to prove the main convergence result, we need additional assumption to ensure some nice stochastic properties of the SAA estimators. The subsequent results lean on the uniform law of large numbers which provides almost sure convergence of the relevant functions. As a consequence, all the subsequent convergence results are stated in a stochastic sense - they hold almost surely.

Assumption 4 *The sample $\xi_1, \xi_2, \ldots$ is i.i.d. and H and its Jacobian ∇H are dominated by P-integrable functions on any compact subset of $\mathbb{R}^n$.*

There is more that one important consequence of Assumptions 1 and 4 ([22]). First, the mapping h is well defined and differentiable. Moreover, the expectation and the gradient are interchangeable $\nabla h(x) = E[\nabla H(x,\xi)]$ and the SAA estimators are unbiased. Furthermore, the uniform law of large numbers implies that the following holds almost surely for any given compact set S

$$\lim_{N\to\infty} \max_{x\in S} \|\nabla h_N(x) - \nabla h(x)\| = 0, \quad \lim_{N\to\infty} \max_{x\in S} \|h_N(x) - h(x)\| = 0. \tag{12}$$

This furthermore implies that the same is true if we substitute h in (12) for θ. It also holds for ϕ with the fixed penalty parameter. We state this within the following technical lemma which is a direct consequence of Theorems 7.48 and 7.52 from [22].

Lemma 2 *Suppose that Assumptions 1 and 4 hold. Then, for any given compact set $S \subset \mathbb{R}^n$ and penalty parameter $\mu > 0$ there holds*

$$\lim_{N\to\infty} \varphi_N^S = 0 \quad a.s., \quad where \quad \varphi_N^S := \max_{x\in S} |\phi_N(x;\mu) - \phi(x;\mu)|.$$

Moreover, for any $\psi \in \{h, \nabla h, \theta, \nabla\theta, \phi(\cdot;\mu), \nabla\phi(\cdot;\mu)\}$ there holds

$$\lim_{x\to x^*, N\to\infty} \psi_N(x) = \psi(x^*) \quad a.s.$$

The subsequent result states that the sequence of penalty parameters tends to infinity, almost surely, if (7) happens only finitely many times. As in the standard quadratic penalty framework, this is needed for ensuring the feasibility. On the other hand, if there are infinitely many iterations such that (7) holds, the feasibility will also be achieved under the additional assumptions stated above, so (7) does not effect the main result.

Theorem 2 *Suppose that Assumptions 1-4 hold and that the sequence* $\{x_k\}_{k\in\mathbb{N}_0}$ *generated by Algorithm 1 is bounded. If there is a finite* q *such that (7) is violated for every* $k \geq q$ *, then*

$$\lim_{k\to\infty} \mu_k = \infty \quad a.s.$$

Proof Suppose that there exist $\bar{q} \geq q$ and $\bar{\mu}$ such that

$$\mu_k = \bar{\mu} \quad \text{for every} \;\; k \geq \bar{q}.$$

Considering Step 6 of Algorithm 1, this means that

$$dm_k > \frac{\alpha_k}{\mu_k^2} \quad \text{for every} \;\; k \geq \bar{q}, \tag{13}$$

i.e.,

$$-g_k^T d_k > \frac{1}{\bar{\mu}^2} > 0 \quad \text{for every} \;\; k \geq \bar{q}. \tag{14}$$

Furthermore, since the sequence of iterates is assumed to be bounded, there exists a compact set S such that $\{x_k\}_{k\in\mathbb{N}_0} \subseteq S$. Moreover, Theorem 1 implies that N_k tends to infinity and thus Lemma 2 implies the existence of the sequence $\varphi_{N_k}^S$ such that

$$|\phi_{N_k}(x_j;\bar{\mu}) - \phi(x_j;\bar{\mu})| \leq \varphi_{N_k}^S \quad \text{for every} \;\; k, j \geq \bar{q}$$

and $\varphi_{N_k}^S$ tends to zero a.s.

In order to prove the statement of this theorem by contradiction, we observe two complementarity cases regarding the step size. First, assume that the step size sequence is bounded away from zero, i.e., there is $\bar{\alpha} > 0$ such that $\alpha_k \geq \bar{\alpha}$ for every $k \geq \bar{q}$. This, together with (13), implies that $dm_k \geq \bar{\alpha}/\bar{\mu}^2 := \bar{d}$ for every $k \geq \bar{q}$ so the line search implies

$$\phi(x_{k+1};\bar{\mu}) \leq \phi_{N_k}(x_{k+1};\bar{\mu}) + \varphi_{N_k}^S \leq \phi_{N_k}(x_k;\bar{\mu}) - \eta dm_k + \varphi_{N_k}^S \leq \phi(x_k;\bar{\mu}) - \eta\bar{d} + 2\varphi_{N_k}^S.$$

Now, since $\varphi_{N_k}^S \to 0$ a.s., there exist $\bar{k} \geq \bar{q}$ and $c > 0$ such that

$$\phi(x_{k+1};\bar{\mu}) \leq \phi(x_k;\bar{\mu}) - c \quad \text{a.s. for every} \quad k \geq \bar{k}.$$

This furthermore implies that $\phi(x_k;\bar{\mu})$ is a.s. unbounded from below which is in contradiction with the Assumption 1.

Now, let us consider the remaining case where $\lim_{k\in K} \alpha_k = 0$ for some infinite subset $K_1 \subseteq \mathbb{N} \bigcap \{\bar{k}, \bar{k}+1, ...\}$. Without loss of generality we can assume that $\alpha_k < 1$

for every $k \in K_1$. This means that the full step is not accepted by the line search and that for every $k \in K_1$ there exists α_k' such that $\alpha_k = \beta \alpha_k'$ and

$$\phi_{N_k}(x_k + \alpha_k' d_k; \bar{\mu}) > \phi_{N_k}(x_k; \bar{\mu}) + \eta \alpha_k' g_k^T d_k.$$

Due to mean value theorem, for every $k \in K_1$ there exists $t_k \in (0,1)$ such that

$$\nabla_x^T \phi_{N_k}(x_k + t_k \alpha_k' d_k; \bar{\mu}) d_k - \eta g_k^T d_k > 0. \tag{15}$$

Since $\{x_k\}$ is bounded, there exist $K_2 \subseteq K_1$ and x^* such that $\lim_{k \in K_2} x_k = x^*$. Moreover, Lemma 2 implies that $\lim_{k \in K_2} g_k = \nabla_x \phi(x^*; \bar{\mu}) := g^*$ a.s., which together with (4) implies that $\{d_k\}_{k \in K_2}$ is bounded a.s. Thus, a.s., there are $K_3 \subseteq K_2$ and d^* such that $\lim_{k \in K_3} d_k = d^*$. Now, using the fact that $\lim_{k \in K_3} \alpha_k' = 0$ and taking a limit over K_3 in (15) we obtain $(g^*)^T d^* (1 - \eta) \geq 0$, that is, $(g^*)^T d^* \geq 0$ a.s. which is in contradiction with (14).

Since both complementarity cases led to contradiction, we conclude that the sequence of penalty parameters is a.s. unbounded and the statement of this theorem follows from the fact that $\{\mu_k\}_{k \in \mathbb{N}}$ is nondecreasing. □

Finally, we prove the main result. Notice that the consequence of the previous theorem is that the sequence of z_t generated in Step 6 of the main algorithm is infinite under the stated conditions. In that case we prove below that every accumulation point of that sequence is stationary for infeasibility measure θ a.s. Moreover, it is a KKT point of (1) if linear independence constraint qualification (LICQ) holds. Therefore, including the case when (7) happens infinitely many times, we obtain the following result.

Theorem 3 *Suppose that Assumptions 1-4 hold and that the sequence $\{x_k\}_{k \in \mathbb{N}_0}$ generated by Algorithm 1 is bounded. Then there exists an accumulation point x^* of the sequence $\{x_k\}_{k \in \mathbb{N}_0}$ which is stationary for θ almost surely. Moreover, if LICQ holds, then x^* is almost surely a KKT point of the original problem* (1).

Proof First, assume that there is an infinite subsequence $J_1 \subseteq \mathbb{N}$ such that (7) holds for every $k \in J_1$. Since $\{x_k\}_{k \in \mathbb{N}_0}$ is assumed to be bounded, there exist $J_2 \subseteq J_1$ and x^* such that $\lim_{k \in J_2} x_k = x^*$. Moreover, Theorem 1 implies that N_k tends to infinity and thus Lemma 2 implies that

$$h(x^*) = \lim_{k \in J_2} h_{N_k}(x_k) = 0 \quad \text{a.s.}$$

so the accumulation point is feasible and consequently $\nabla \theta(x^*) = 0$ a.s. Furthermore, since for every $k \in J_2$ there holds

$$0 = g_k = \nabla f(x_k) + 2\mu_k \nabla^T h_{N_k}(x_k) h_{N_k}(x_k) = \nabla f(x_k),$$

taking the limit over J_2 we obtain that $\nabla f(x^*) = 0$ and thus x^* is obviously a KKT point of (1).

Now, let us consider the remaining case where (7) happens at most finitely many times. In this case it also holds that $N_k \to \infty$. Moreover, Theorem 2 implies that μ_k

tends to infinity a.s. and, consequently, the subsequence $\{z_t\}$ of $\{x_k\}_{k\in\mathbb{N}}$ is a.s. infinite. We will prove that every accumulation point of the sequence $\{z_t\}_{t\in\mathbb{N}}$ is stationary for θ almost surely. In order to do that, let x^* be an arbitrary accumulation point of the sequence $\{z_t\}$, i.e.,

$$\lim_{k\in K_1} x_k = x^*$$

for some infinite $K_1 \subseteq \mathbb{N}$. Moreover, Step 6 of Algorithm 1 implies that $dm_k \leq \alpha_k/\mu_k^2$ holds for every $k \in K_1$. According to the definition of dm_k, this is equivalent to

$$-g_k^T d_k \leq \frac{1}{\mu_k^2} \quad \text{for every } \; k \in K_1.$$

So, taking a limit over K_1 and using (3) we conclude that

$$\lim_{k\in K_1} g_k = 0. \tag{16}$$

Moreover, since $g_k = \nabla f(x_k) + \mu_k \nabla \theta_{N_k}(x_k)$, we obtain

$$\|\nabla \theta_{N_k}(x_k)\| \leq \frac{1}{\mu_k}(\|g_k\| + \|\nabla f(x_k)\|)$$

and taking a limit yields $\lim_{k\in K_1} \|\nabla \theta_{N_k}(x_k)\| = 0$. Furthermore, Lemma 2 implies that $\lim_{k\in K_1} \nabla \theta_{N_k}(x_k) = \nabla \theta(x^*)$ a.s. so

$$\nabla \theta(x^*) = 0 \quad \text{a.s.}$$

Moreover, suppose that LICQ holds. Then $\nabla h(x^*)$ has a full rank and since $\nabla \theta(x^*) = 2\nabla^T h(x^*)h(x^*)$, we conclude that $h(x^*) = 0$ a.s. Furthermore, let us define

$$\lambda_k := 2\mu_k h_{N_k}(x_k).$$

Then it holds

$$\nabla^T h_{N_k}(x_k)\lambda_k = g_k - \nabla f(x_k). \tag{17}$$

Moreover, Lemma 2 implies that $\lim_{k\in K_1} \nabla h_{N_k}(x_k) = \nabla h(x^*)$ a.s. so $\nabla h_{N_k}(x_k)$ has a full rank a.s. for all $k \in K_1$ large enough and we can use the following representation

$$\lambda_k = \left(\nabla h_{N_k}(x_k)\nabla^T h_{N_k}(x_k)\right)^{-1} \nabla h_{N_k}(x_k)\left(g_k - \nabla f(x_k)\right).$$

Now, taking a limit over K_1 and using (16) we obtain

$$\lim_{k\in K_1} \lambda_k = -\left(\nabla h(x^*)\nabla^T h(x^*)\right)^{-1} \nabla h(x^*)\nabla f(x^*) \quad \text{a.s.} \tag{18}$$

Hence, denoting the right hand side of (18) by λ^* and taking a limit over K_1 in (17) we conclude that

$$\nabla f(x^*) + \nabla^T h(x^*)\lambda^* = 0 \quad \text{a.s.},$$

which together with the feasibility implies that x^* is a KKT point of the problem (1) a.s. This completes the proof. □

Remark 2 Notice that the same result can be stated if we assume that $H(\cdot,\xi) \in C^1(\mathbb{R}^n)$ for *almost* every ξ instead of "every ξ" as stated within the Assumption 1. However, the intermediate results such as one stated in Theorem 1 no longer hold since the sample size tends to infinity almost surely in that case. Moreover, the gradient g_k is well defined almost surely and the same is true for the proposed algorithm. Although the concept alters, the final result still holds under this relaxation.

4 Numerical results

In this section we present the numerical results obtained on 13 equality constrained optimization problems from Hock and Schittkowski [7] (problems 6, 27, 28, 42, 46-52, 61 and 79). All the problems have unique solutions and the objective functions bounded from below. We randomize the original constraints $c(x)$ as follows:

$$H(x,\xi) = c(\xi x)$$

where ξ follows the normal distribution $\mathcal{N}(1,1)$.

The proposed sample size update (VSS) is compared with the heuristic scheme (HEUR) where

$$N_{k+1} = \lceil \min\{1.1N_k\} \rceil,$$

as in [6,12,14,17]. The penalty parameter is updated as in Algorithm 1. We performed 10 runs of each method for every considered problem. The runs are conducted in Matlab. We used the built-in function *randn* to generate the random samples. Within each run, VSS and HEUR are tested with the same sample realization. Moreover, all the remaining common parameters are identical. Both methods use the BFGS search direction where the gradient difference is calculated as

$$y_k = \nabla_x \phi_{N_{k+1}}(x_{k+1};\mu_k) - \nabla_x \phi_{N_k}(x_k;\mu_k)$$

and $s_k = x_{k+1} - x_k$. The descent property of d_k is provided by using the safeguard - starting with the identity matrix, the BFGS matrix is updated only if $y_k^T s_k > 10^{-6}$. The increase factor of the penalty parameter is $\rho = 1.5$ and the initial value is $\mu_0 = 1$. The initial iteration points x_0 are as in Hock and Schittkowski [7] and the starting sample size is $N_0 = 3$. The line search parameters are $\beta = 0.5$ and $\eta = 10^{-4}$.

We specify VSS algorithm as follows. Function γ used in Step 2 of Algorithm 2 is defined as

$$\gamma(N) = exp(-1/N),$$

which satisfies Assumption 3. On the other hand, the SAA error measure is set to

$$e(x;N) = \frac{1.96\hat{\sigma}_N(x)}{\sqrt{N}},$$

where $\hat{\sigma}_N^2(x)$ represents the measure of the sample variance

$$\hat{\sigma}_N^2(x) = \frac{1}{N-1}\sum_{i=1}^{N} \|H(x,\xi_i) - h_N(x)\|^2$$

and 1.96 stands for approximation of the 0.975 quantile of the standard normal distribution. This choice corresponds to the confidence interval width measure for $h_N(x)$ estimator. Although it is not bounded away from zero in general, we use this type of measure as it is commonly used as a less conservative variant of the log bound, [1, 11–15,18], and it takes into account the current point x. Assumption 2 can be satisfied simply by adding $1/N$, for instance, to the measure e defined above. In our experience, the less conservative case works well in practice.

The comparison of the procedures is based on the number of evaluations of the constraints function H and the components of its Jacobian ∇H denoted by FEV. Since the algorithms do not have a stopping criterion, the runs are stopped when 10^6 of FEVs is reached. For problems with linear constraints, the transformed problems of the form (1) are equivalent to the original ones from [7] and thus we can use the solutions provided therein. Solutions of the problems with nonlinear constraints are obtained by finding the analytical form of the relevant mathematical expectation and solving the obtained deterministic problem approximately with a high precision by applying the built-in Matlab function *fmincon*. Knowing the optimal or nearly-optimal solutions gives us a chance to have more insights in the performance of the proposed scheme.

Let us denote the obtained solution of the considered problem by x^*. Our aim is to show that VSS scheme reaches the vicinity of the solution with less efforts (i.e., FEVs) than the HEUR scheme. In order to do that, we observe different precisions τ of the solution approximations. In other words, we seek for a number of FEVs which provides a τ-optimal solution for different values of τ. More precisely, we track the solution estimate x_k and the relevant $FEV_k = FEV(x_k)$ at every iteration for both methods and, considering 10 runs of each problem, we find the empirical probability (denoted by P_E) that the method is not worse than the other one. In order to make a fair comparison, if the same number of FEVs is needed we assume that both VSS and HEUR are the winners and we present empirical probabilities for both schemes:

$$p_\tau^{VSS} := P_E\left\{\min_k\{FEV_k^{VSS} : \|x_k^{VSS} - x^*\| \le \tau\} \le \min_l\{FEV_l^{HEUR} : \|x_l^{HEUR} - x^*\| \le \tau\}\right\}$$

and

$$p_\tau^{HEUR} := P_E\left\{\min_k\{FEV_k^{HEUR} : \|x_k^{HEUR} - x^*\| \le \tau\} \le \min_l\{FEV_l^{VSS} : \|x_l^{VSS} - x^*\| \le \tau\}\right\}$$

for different levels of precision $\tau \in (0,2]$. The results presented in Figure 1 show that the proposed scheme outperforms the observed heuristic approach. Moreover, if less precise approximation of the solution is needed, the advantage is even bigger. Of course, these conclusions are based only on the tested problems experience. Moreover, for any given problem, one can always find a heuristic that outperforms VSS, but we believe that the advantage of the proposed scheme lies in its adaptive nature. We present a typical behavior of the sequences $\{N_k\}$ and $\{\mu_k\}$ in Figure 2.

5 Conclusions

Focus of this research is on the problems with equality constraints in the form of mathematical expectation. The proposed method incorporates the sample average approximations with adaptive sample size scheduling into the quadratic penalty framework. Sample is assumed to be cumulative and the sample size update is based on balancing of two types of errors - the SAA error and the measure of optimality.

The proposed algorithm represents nontrivial generalization of the method that copes with the finite sample and solves only an approximate problem. Instead of that,

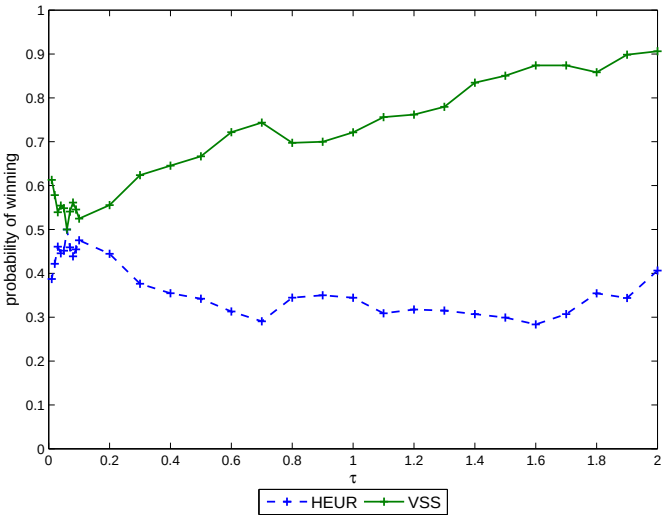

Fig. 1 Empirical probabilies of winning p_τ^{VSS} and p_τ^{HEUR} for different levels of accuracy τ

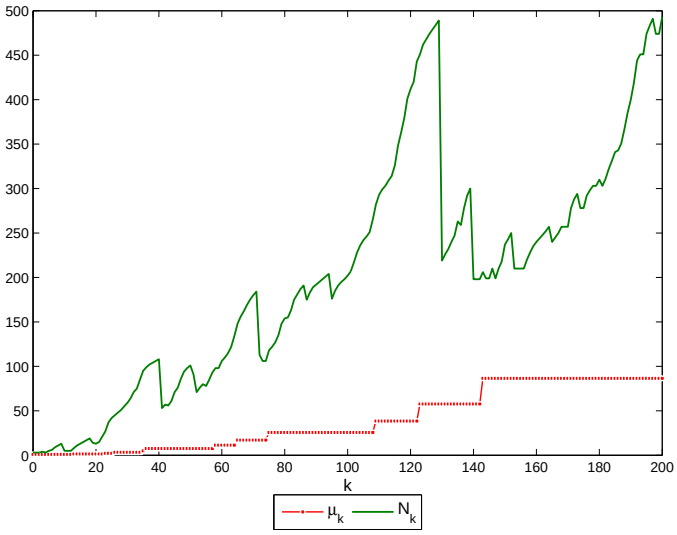

Fig. 2 Typical behavior of the sample size and penalty parameter sequences

we aim to solve the original problem and thus the sample size needs to tend to infinity. In Theorem 1 we prove that the proposed mechanism ensures that if the objective function is bounded from below and all the considered functions are continuously differentiable. We also prove (Theorem 2) that the penalty parameter tends to infinity almost surely if the function under the mathematical expectation and its Jacobian are dominated by P-integrable functions on any compact set. Since the penalty parameter is also updated adaptively, this result requests nontrivial analysis. The remaining

assumptions can be considered as a guidance for choosing d_k, $e(x;N)$ and $\gamma(N)$ and they are easy to satisfy. Finally, under the stated assumptions, we prove (Theorem 3) that there is a subsequence of iterates which converges to a KKT point almost surely if the sequence of iterates is bounded and LICQ hold.

Numerical study is performed on a class of test problems with the objective function bounded from below. We considered both linear and nonlinear constrained cases. The results show that the proposed adaptive scheme outperforms the considered heuristic with predetermined sample scheduling in terms of optimization costs measured throughout the number of function evaluations. More precisely, by employing the empirical probability, we show that the proposed scheduling attains a near-optimal solution with less efforts than the heuristic.

Although the number of tested problems is modest, the difference in favor of the proposed method is rather significant and we believe that the adaptive scheme has a potential to be highly competitive in general. On the other hand, tuning the functions $e(x;N)$ and $\gamma(N)$ for specific classes of problems would surely provide even better performance and this will be the topic of our future research.

Acknowledgements We are grateful to the anonymous referee whose comments and suggestions helped us to improve the quality of this paper.

References

1. Bastin, F.: Trust-Region Algorithms for Nonlinear Stochastic Programming and Mixed Logit Models, PhD thesis. University of Namur, Belgium (2004)
2. Bastin, F., Cirillo, C., Toint, P.L.: An adaptive Monte Carlo algorithm for computing mixed logit estimators, Comput. Manag. Sci. 3(1), 55-79 (2006)
3. Bastin, F., Cirillo, C., Toint, P.L.: Convergence theory for nonconvex stochastic programming with an application to mixed logit, Math. Program. 108(2-3), 207–234 (2006)
4. Birge, J. R., Louveaux, F.: Introduction to Stochastic Programming. Springer Series in Operations Research and Financial Engineering, Springer Science+Business Media, LLC, New York (2011)
5. Deng, G., Ferris, M.C.: Variable-number sample path optimization, Math. Program. 117(12), 81-109 (2009)
6. Friedlander, M.P., Schmidt, M.: Hybrid deterministic-stochastic methods for data fitting, SIAM. J. Sci. Comput. 34(3), 1380-1405 (2012)
7. Hock, W., Schittkowski, K.: Test Examples for Nonilnear Programming Codes, Lecture Notes in Economics and Mathematical Systems 187, Springer (1981)
8. Homem-de-Mello, T.: Variable-Sample Methods for Stochastic Optimization, ACM Trans. Model. Comput. Simul. 13(2), 108–133 (2003)
9. Homem-de-Mello, T.: On rates of convergence for stochastic optimization problems under noninde-pendent and identically distributed sampling, SIAM J. Optim. 19(2), 524–551 (2008)
10. Kleywegt, A., Shapiro, A., Homem-de-Mello, T.: The sample average approximation method for stochastic discrete optimization, SIAM J. Optim. 12(2), 479-502 (2001)
11. Krejić, N., Krklec, N.: Line search methods with variable sample size for unconstrained optimization, J. Comput. Appl. Math. 245, 213-231 (2013)
12. Krejić, N., Krklec Jerinkić, N.: Nonmonotone line search methods with variable sample size, Numer. Algorithms 68(4), 711-739 (2015)
13. Krejić, N., Krklec Jerinkić, N.: Spectral projected gradient method for stochastic optimization, J Global Optim 73(1), 59-81 (2019)
14. Krejić, N., Krklec Jerinkić, N., Rožnjik, A.: Variable sample size method for equality constrained optimization problems, Optim. Lett. 12(3), 485–497 (2018)
15. Linderoth, J., Shapiro, A., Wright, S.: The empirical behavior of sampling methods for stochastic programming, Ann Oper Res 142, 215–241 (2006)

16. Nocedal, J., Wright, S. J.: Numerical Optimization, Springer Series in Operations Research, Springer, New York, (2006)
17. Pasupathy, R.: On choosing parameters in retrospective-approximation algorithms for stochastic root finding and simulation optimization, Oper. Res. 58(4), 889-901 (2010)
18. Polak, E., Royset, J.O.: Eficient sample sizes in stochastic nonlinear programing, J. Comput. Appl. Math. 217(2), 301-310 (2008)
19. Royset, J. O., Szechtman, R.: Optimal Budget Allocation for Sample Average Approximation, Oper. Res. 61(3), 777–790 (2013)
20. Santoso, T., Ahmed, S., Goetschalckx, M., Shapiro, A.: A stochastic programming approach for supply chain network design under uncertainty, Eur J Oper Res 167, 96–115 (2005)
21. Shapiro, A.: Monte Carlo Sampling Methods. In: Stochastic programming, Handbook in Operations Research and Management Science 10, Elsevier, 353–425 (2003)
22. Shapiro, A., Dentcheva, D., Ruszczynski, A.: Lectures on Stochastic Programming: Modeling and Theory. MPS-SIAM Series on Optimization (2009)
23. Wallace, S.W., Ziemba, W.T. (eds.): Applications of Stochastic Programming. SIAM, Philadelphia (2005)
24. Wang, W., Ahmed, S.: Sample average approximation of expected value constrained stochastic programs, Oper. Res. Lett. 36(5), 515-519 (2008)