Stochastic Successive Convex Approximation for General Stochastic Optimization Problems with Applications in Wireless Communications

Chencheng Ye
Shanghai Jiao Tong University

Ying Cui
Shanghai Jiao Tong University

Abstract—One key challenge for solving a general stochastic optimization problem with expectations in the objective and constraint functions using ordinary stochastic iterative methods lies in the infeasibility issue caused by the randomness over iterates. This letter aims to address this main challenge. First, we obtain an equivalent stochastic optimization problem which is to minimize the weighted sum of the original objective and the penalty for violating the original constraints. Then, we propose a stochastic successive convex approximation (SSCA) method to obtain a stationary point of the original stochastic optimization problem. Using similar techniques, we propose a parallel SSCA method to obtain a stationary point of a special case of the general stochastic optimization problem which has decoupled constraint functions. We also provide application examples of the proposed methods in power control for interference networks. The proposed SSCA and parallel SSCA methods achieve empirically higher convergence rates and lower computational complexities than existing ones, benefiting from the elegant way of balancing the objective minimization and constraint satisfaction over random iterates.

I. INTRODUCTION

Stochastic optimization problems refer to optimization problems which involve random variables. They are of broad interest, with applications arising in wireless communications, business analytics, manufacturing, finance, etc. In general, stochastic optimization problems with expectations in possibly nonconvex objective functions can be classified into three categories, namely, unconstrained stochastic problems, stochastic optimization problems with deterministic constraints, and stochastic optimization problems with expectations in constraint functions. Without loss of generality, in the following, we restrict our attention to stochastic minimization problems.

In [1], a stochastic gradient method is proposed to obtain a stationary point of an unconstrained stochastic optimization problem. In [2–4], the stochastic gradient averaging method [2], stochastic majorization-minimization (MM) method [3] and stochastic successive convex approximation (SSCA) method [4] are proposed to obtain a stationary point of a stochastic optimization problem with deterministic convex constraints. Stochastic optimization problems with expectations in constraint functions are more challenging, as the stochastic nature of the constraint functions may cause infeasibility at each iteration of an ordinary stochastic iterative method. In [5], an SSCA method is proposed to directly tackle a general stochastic optimization problem with expectations in the constraint functions, for the first time. Specifically, at each iteration, an approximate convex problem is solved to minimize the objective; if it is infeasible, another approximate convex problem is then solved to minimize the penalty for violating the constraints. Using similar techniques, a parallel SSCA method is proposed for a special case of the general stochastic optimization problem with decoupled constraint functions. Leveraging two types of approximate problems with different goals at each iteration can deal with the infeasibility issue, but may lead to decrease of convergence rate and increment of computational complexity.

In this letter, we shall address the above issue. As in [5], we consider a general stochastic optimization problem with expectations in the constraint functions. First, we obtain an equivalent stochastic optimization problem whose objective function is the weighted sum of the original objective and the penalty for violating the original constraints. Then, we propose an SSCA method that involves solving an approximate convex optimization problem which is always feasible at each iteration. Moreover, we show that the proposed SSCA method converges to a stationary point of the equivalent stochastic optimization problem, which is also a stationary point of the original stochastic optimization problem under certain conditions. Using similar techniques, we propose a parallel SSCA method to obtain a stationary point of a special case of the aforementioned general stochastic optimization problem which has decoupled constraint functions. As application examples, we consider the optimal power allocation to maximize the ergodic sum-rate under the coupled and decoupled individual ergodic rate constraints, respectively, and illustrate how to apply the proposed SSCA and parallel SSCA methods to obtain their respective stationary points. Numerical results show that the proposed SSCA and parallel SSCA methods have higher convergence rates and lower computational complexities than those in [5]. The substantial gains derive from the effective balance of the minimization of the original objective and the satisfaction of the original constraints over random iterates.

II. GENERAL STOCHASTIC OPTIMIZATION

In this section, we consider a general stochastic optimization problem with expectations in both the objective and constraint functions that are possibly nonconvex.
Problem 1 (General Stochastic Optimization Problem):

\[
\min_x f_0(x) \triangleq \mathbb{E} [g_0(x, \xi)] \\
\text{s.t. } f_i(x) \triangleq \mathbb{E} [g_i(x, \xi)] \leq 0, \quad i = 1, \ldots, m, \\
x \in \mathcal{X},
\]

where \( x \triangleq (x_1, \ldots, x_n) \) is the optimization variable, \( \mathcal{X} \subseteq \mathbb{R}^n \), \( \xi \) is a random vector defined on the probability space \((\Omega, \mathcal{F}, \mathbb{P})\) with \( \Omega \) being the sample space, \( \mathcal{F} \) being the \( \sigma \)-algebra generated by subsets of \( \Omega \), and \( \mathbb{P} \) being a probability measure defined on \( \mathcal{F} \), and functions \( g_i : \mathcal{X} \times \Omega \rightarrow \mathbb{R}, i = 0, \ldots, m \) are possibly nonconvex.

Assumption 1 (Assumptions on Problem 1 [7, 8, 9]):

1) \( \mathcal{X} \) is compact and convex;
2) For any given \( \xi \), each \( g_i(x, \xi) \) is continuously differentiable on \( \mathcal{X} \), and its gradient is Lipschitz continuous.

Problem 1 is very challenging, and is not well studied. First, motivated by the feasible point pursuit method in [6], we transform Problem 1 to the following stochastic optimization problem whose objective function is the weighted sum of the original objective function and the slack penalty term.

Problem 2 (Equivalent Problem of Problem 1):

\[
\min_{x, s} f_0(x) + \rho \sum_{i=1}^{m} s_i \\
\text{s.t. } f_i(x) \leq s_i, \quad i = 1, \ldots, m, \\
s_i \geq 0, \quad i = 1, \ldots, m,
\]

where \( s \triangleq (s_i)_{i=1}^{m} \) are slack variables and \( \rho > 0 \) is a penalty parameter that trades off the original objective function and the slack penalty term.

Note that Problem 2 is always feasible. The relationship between Problem 1 and Problem 2 is summarized below.

Lemma 1 (Equivalence between Problem 1 and Problem 2): If Problem 1 is feasible and Assumption 1 is satisfied, then there exists \( \rho_0 \geq 0 \) such that for all \( \rho \geq \rho_0 \), Problem 2 and Problem 1 have the same optimal value.

Proof: As \( g_0(x, \xi) \) is continuous and \( \mathcal{X} \) is compact, \( g_0(x, \xi) \) is bounded on \( \mathcal{X} \). Thus, the optimal value of Problem 1 is bounded if Problem 1 is feasible. Therefore, by Lemma 1, we know that there exists \( \rho_0 \geq 0 \) such that for all \( \rho \geq \rho_0 \), Problem 2 and Problem 1 have the same optimal value. \( \square \)

Based on Lemma 1, we now focus on solving Problem 2. Like Problem 1, Problem 2 is a stochastic optimization problem with possibly nonconvex objective and constraint functions. In the following, we propose an effective SSCA method to obtain a stationary point of Problem 2 using the SCMA technique [3]. Later, we shall show that under certain conditions, a stationary point of Problem 2 is also a stationary point of Problem 1.

Specifically, at iteration \( t \), we solve the following approximate convex optimization problem of Problem 2

Problem 3 (Approximate Convex Optimization Problem in \( t \)-th Iteration):

\[
\min_{x, s} \tilde{f}_0^t(x) + \rho \sum_{i=1}^{m} s_i \\
\text{s.t. } \tilde{f}_i(x) \leq s_i, \quad i = 1, \ldots, m,
\]

where \( \tilde{f}_i^t(x), i = 0, \ldots, m \) are convex surrogate functions of \( f_i(x), i = 0, \ldots, m \). Let \((x^t, s^t)\) denote an optimal solution of Problem 3.

Problem 3 is a convex optimization problem which is always feasible and can be solved with conventional convex optimization techniques. Given \( x^t \), we update \( x^t \) according to:

\[
x^{t+1} = (1 - \gamma^t)x^t - \gamma^t x^t, \quad t = 1, 2, \ldots,
\]

where \( \gamma^t \) is a positive diminishing stepsize satisfying:

\[
\gamma^t = 0, \quad \lim_{t \rightarrow \infty} \gamma^t = 0, \quad \sum_{t=1}^{\infty} \gamma^t = \infty, \quad \sum_{t=1}^{\infty} (\gamma^t)^2 < \infty.
\]

The details are summarized in Alg. 1. To ensure the convergence of Alg. 1, the surrogate functions \( f_i^t(\cdot), i = 0, \ldots, m \) should satisfy the following assumptions.

Assumption 2 (Assumptions on \( f_i^t(\cdot) \) [4, 5]):

1) Each \( f_i^t(x) \) is uniformly strongly convex on \( \mathcal{X} \);
2) Each \( f_i^t(x) \) is Lipschitz continuous on \( \mathcal{X} \), and for any \( x \in \mathcal{X}, \lim_{t \rightarrow \infty} \sup_{x \in \mathcal{X}} f_i^t(x) - f_i^t(x) \leq B\|x^t - x^i\|_2 \), for some constant \( B > 0 \);
3) Each \( \{\nabla_x f_i^t(x) : t = 0, 1, \ldots\} \) is uniformly bounded;
4) \( \lim_{t \rightarrow \infty} \|\nabla f_i^t(x) - \nabla f_i^t(x)\|_2 = 0 \) and \( \lim_{t \rightarrow \infty} \|\nabla f_i^t(x) - \nabla f_i^t(x)\|_2 = 0 \).

A common example of surrogate functions is [4, 5]:

\[
\tilde{f}_i^t(x) = (1 - \omega^t) f_i^t(x) + \omega^t g_i(x, x_i^t - x_i^t), \quad i = 0, \ldots, m, \quad t = 1, 2, \ldots,
\]

where \( f_i^0(x) = 0 \) for all \( x \in \mathcal{X} \), \( \omega^t \) is a positive diminishing stepsize satisfying:

\[
\omega^t > 0, \quad \lim_{t \rightarrow \infty} \omega^t = 0, \quad \sum_{t=1}^{\infty} \omega^t = \infty, \quad \sum_{t=1}^{\infty} (\omega^t)^2 < \infty, \quad \lim_{t \rightarrow \infty} \frac{\gamma^t}{\omega^t} = 0.
\]

and \( g_i(x, x_i^t, x_{i-1}^t) \) is a convex approximation of \( g_i(x, x_{i-1}^t) \) around \( x_{i-1}^t \) satisfying:

\[
\tilde{g}_i(x, x, \xi) = g_i(x, \xi) \quad \text{and} \quad \nabla_x \tilde{g}_i(x, x, \xi) = \nabla g_i(x, \xi) \quad \text{for all} \quad x \in \mathcal{X} \quad \text{and} \quad \xi \in \Omega.
\]

\( \tilde{g}_i(x, y, \xi) \) is strongly convex in \( x \) for all \( y \in \mathcal{X} \) and \( \xi \in \Omega \):

\( \tilde{g}_i(x, y, \xi) \) is Lipschitz continuous in both \( x \) and \( y \) for all \( \xi \in \Omega \). It has been shown in [5, Proposition 1] that if the stepizes \( \{\gamma^t\} \) and \( \{\omega^t\} \) satisfy (7) and (9), respectively, then the surrogate functions given by (17) satisfy Assumption 2.

Finally, we show the convergence of Alg. 1.

Theorem 1 (Convergence of Alg. 1): Suppose Assumption 1 and Assumption 2 are satisfied. Then \( \{(x^t, s^t)\} \) generated by Alg. 1 has a limit point, denote by \((x^*, s^*)\), and the following statements hold:

1) \((x^*, s^*)\) is a stationary point of Problem 2
2) If \( s^* = 0 \), then \( x^* \) is a stationary point of Problem 1.
Algorithm 1 SSCA

1: initialization: Set \( t = 1 \), and choose any \( x^0 \in \mathcal{X} \).
2: repeat
3: Obtain \((\hat{x}^t, s^t)\) by solving Problem 5 with conventional convex optimization techniques, and update \( x^t \) according to (6).
4: Set \( t = t + 1 \).
5: until Some convergence criteria is met.

Proof: By Assumption 1, we know that \( x^t \) is bounded. As \((\hat{x}^t, s^t)\) is an optimal solution of Problem 5, it can be easily shown that \( s^t_i = f^*_i(\hat{x}^t) \), \( i = 1, \ldots, m \). By Assumption 2 and Assumption 2.2, we know that \( \bar{f}^*_i(\hat{x}^t) \) is bounded, which implies that \( s^t_i \) is bounded. Note that Assumption 1.2, Assumption 2 and Assumption 3 in [5] readily follow Assumption 1.2 and Assumption 2 in this letter, and Assumption 1.1 in [5] is used to prove the boundedness of \( \bar{f}^*_i(\hat{x}^t) + s^t_i \). Therefore, following the proof of [5] Theorem 1, we can show the first statement. In addition, it can be easily shown that when \( s = 0 \), the KKT conditions of Problem 2 imply those of Problem 1. Therefore, we can show the second statement.

Note that we can run Alg. 1 multiple times, each with a random initial point \( x^0 \in \mathcal{X} \), until a stationary point \((x^*, s^*)\) of Problem 2 with \( s^* = 0 \), i.e., a stationary point of Problem 1 is obtained.

III. STOCHASTIC OPTIMIZATION WITH DECOUPLED CONSTRAINTS

In this section, we consider a special case of the general stochastic optimization problem, which has decoupled constraint functions involving expectations.

Problem 4 (Stochastic Optimization Problem with Decoupled Constraints):

\[
\begin{align*}
\min_{x} & \ f_0(x) \triangleq \mathbb{E}[g_0(x, \xi)] \\
\text{s.t.} & \ \ f_{k,i}(x_k) \triangleq \mathbb{E}[g_{k,i}(x_k, \xi)] \leq 0, \ k = 1, \ldots, K, \ i = 1, \ldots, m_k, \\
& \ x_k \in \mathcal{X}_k, \ k = 1, \ldots, K,
\end{align*}
\]

where the optimization variable can be partitioned into \( K \) blocks, i.e., \( x \triangleq (x_k)_{k=1}^K \), with \( x_k \) being the variable for the \( k \)-th block, \( \mathcal{X}_k \subseteq \mathbb{R}^{n_k}, k = 1, \ldots, K, \mathcal{X}_k \) is a random vector, and functions \( g_0 : \mathcal{X} \times \Omega \rightarrow \mathbb{R} \) and \( g_{k,i} : \mathcal{X}_k \times \Omega \rightarrow \mathbb{R} \) are possibly nonconvex.

Assumption 3 (Assumptions on Problem 4 [5]):

1) \( \mathcal{X} \) is compact and convex;
2) For any given \( \xi \), \( g_0(x, \xi) \) and each \( g_{k,i}(x_k, \xi) \) are continuously differentiable on \( \mathcal{X} \) and \( \mathcal{X}_k \), respectively, and their gradients are Lipschitz continuous.

Note that the constraints of Problem 4 can be separated into \( K \) groups with the \( k \)-th group of constraints depending on the \( k \)-th block \( x_k \). Similarly, we transform Problem 4 to the following stochastic optimization problem.

Problem 5 (Equivalent Problem of Problem 4):

\[
\begin{align*}
\min_{x,*} & \ f_0(x) + \rho \sum_{k=1}^K \sum_{i=1}^{m_k} s_{k,i} \\
\text{s.t.} & \ f_{k,i}(x_k) \leq s_{k,i}, \ k = 1, \ldots, K, \ i = 1, \ldots, m_k, \\
& \ s_{k,i} \geq 0, \ k = 1, \ldots, K, \ i = 1, \ldots, m_k
\end{align*}
\]

where \( s \triangleq (s_{k,i})_{i=1}^{m_k}, \ k = 1, \ldots, K \) are slack variables and \( \rho > 0 \) is the penalty parameter.

Note that Problem 5 is always feasible. The relationship between Problem 4 and Problem 5 is summarized below.

Lemma 2 (Equivalence between Problem 2 and Problem 5): If Problem 4 is feasible and Assumption 3 is satisfied, then there exists \( \rho_0 \geq 0 \) such that for all \( \rho \geq \rho_0 \), Problem 5 and Problem 4 have the same optimal value.

Proof: The proof is similar to that of Lemma 1 and is omitted due to page limitation.

Based on Lemma 2 we now focus on solving Problem 5. In the following, we propose an effective parallel SSCA method to obtain a stationary point of Problem 5 using the parallel SSCA technique [4]. Similarly, we shall show that under certain conditions, a stationary point of Problem 5 is also a stationary point of Problem 4.

Specifically, at iteration \( t \), we solve the following \( K \) approximate convex optimization problems of Problem 5 one for each block.

Problem 6 (Approximate Convex Optimization Problem for \( k \)-th Block in \( t \)-th Iteration):

\[
\begin{align*}
\min_{x_k, s_k} & \ f_{k,0}^t(x_k) + \rho \sum_{i=1}^{m_k} s_{k,i} \\
\text{s.t.} & \ f_{k,i}^t(x_k) \leq s_{k,i}, \ i = 1, \ldots, m_k, \\
& \ s_{k,i} \geq 0, \ i = 1, \ldots, m_k
\end{align*}
\]

where \( s_k \triangleq (s_{k,i})_{i=1}^{m_k} \) are slack variables and \( f_{k,0}^t(x_k), f_{k,i}^t(x_k), i = 1, \ldots, m_k \) are convex surrogate functions of \( f_0(x_k), f_{k,i}(x_k), i = 1, \ldots, m_k \). Let \((x_k^t, s_k^t)\) denote an optimal solution of Problem 6.

The \( K \) approximate problems can be solved in a distributed and parallel manner using conventional convex optimization techniques [4]. Denote \( x^t_k \triangleq (x_k^t)_{k=1}^K \). The details are summarized in Alg. 2. To ensure the convergence of Alg. 2, the surrogate functions \( f_{k,0}^t(x_k), f_{k,i}^t(x_k), i = 1, \ldots, m_k \) should satisfy the following assumptions.

Assumption 4 (Assumptions on \( f_{k,0}^t(x_k), f_{k,i}^t(x_k) \) [4], [5]):

1) Each \( f_{k,0}^t(x_k) \) is uniformly strongly convex on \( \mathcal{X}_k \);
2) Each \( f_{k,i}^t(x_k) \) is Lipschitz continuous on \( \mathcal{X}_k \), and for any \( x_k \in \mathcal{X}_k \), \( \lim_{t_1, t_2 \rightarrow \infty} f_{k,i}^t(x_k) - f_{k,i}^0(x_k) \leq B \| x_k^1 - x_k^0 \| \), for some constant \( B \geq 0 \);
3) Each \( \{\nabla^2_{xx_{k,i}} f_{k,i}^t(x_k) : t = 0, 1, \ldots\} \) is uniformly bounded;
4) \( \lim_{t \rightarrow \infty} f_{k,i}^t(x_k^t) = f_{k,i}^0(x_k^t) = 0 \) and \( \lim_{t \rightarrow \infty} \| \nabla_{x_{k,i}} f_{k,i}^t(x_k^t) \| = 0 \).

Note that the surrogate objective function of the parallel SSCA method in this letter is more general than that in [5], and can exploit block-wise structures of the objective function.
Algorithm 2 Parallel SSCA

1. Initialization: Set $t = 1$, and choose any $x^0 \in \mathcal{X}$.
2. repeat
3. Obtain $(\tilde{x}_i, s^t_i)$ by solving Problem 6 with conventional convex optimization techniques, for $k = 1, \ldots, K$, and update $x^t$ according to (6).
4. Set $t = t + 1$.
5. until Some convergence criteria is met.

A common example of surrogate functions is given as follows [4, 5]:

\[
\tilde{f}_{k,i}(x_k) = (1 - \omega^t) \tilde{f}_{k,i}^{t-1}(x_k) + \omega^t \tilde{g}_{k,i}(x_k, x^{t-1}, \xi^t),
\]

where $\tilde{f}_{k,i}(x_k) = 0$ for all $x_k \in \mathcal{X}_k$, $\omega^t$ is a positive diminishing stepsize satisfying (9), and $\tilde{g}_{k,i}(x_k, x^{t-1})$ is a convex approximation of $g_{k,i}(x_{k-1}, x^{t-1})$ around $x_{k-1}$ satisfying: $\tilde{g}_{k,i}(x_k, x, \xi) = g_{k,i}(x, \xi)$ and $\nabla_x \tilde{g}_{k,i}(x_k, x, \xi) = \nabla_x g_{k,i}(x, \xi)$, for all $x \in \mathcal{X}$ and $\xi \in \Omega$; $\tilde{g}_{k,i}(x_{k-1}, y, \xi)$ is strongly convex in $x_k$ for all $y \in \mathcal{X}$ and $\xi \in \Omega$; $\tilde{g}_{k,i}(x_k, y, \xi)$ is Lipschitz continuous in both $x$ and $y$ for all $\xi \in \Omega$. Similarly, by [4] Proposition 1, the surrogate functions given by (17) satisfy Assumption 4.

Finally, we show the convergence of Alg. 2.

Theorem 2 (Convergence of Alg. 2). Suppose Assumption 4 and Assumption 5 are satisfied. Then \{$(x^t, s^t)$\} generated by Alg. 2 has a limit point, denote by $(x^\star, s^\star)$, and the following statements hold:

1) $(x^\star, s^\star)$ is a stationary point of Problem 5.
2) If $s^\star = 0$, then $x^\star$ is a stationary point of Problem 4.

Proof: Similarly to the proof of Theorem 1, we can show that $s^\star$ is bounded. Note that Assumption b, Assumption c in [4] readily follow Assumption 1.2 and Assumption 2 in this letter, and Assumption a in [4] is used to prove the boundedness of $f^t(x^t) + s^t$, where $x^t \triangleq [x^t_1]_{k=1}^K$. Thus, following the proof of [4] Theorem 1, we can show Lemma 4 in [5]. Then, following the proof of [5] Theorem 1, we can show the KKT conditions of Problem 5 hold. Therefore, we can show the first statement. Similarly, when $s = 0$, the KKT conditions of Problem 5 imply those of Problem 4. Therefore, we can show the second statement. 

Similarly, we can run Alg. 2 multiple times, each with a random initial point $x^0 \in \mathcal{X}$, until a stationary point of Problem 4 is obtained.

IV. APPLICATION EXAMPLES IN INTERFERENCE NETWORKS

Consider a $K$-pair frequency-selective interference channel. Each pair includes one single-antenna transmitter and one single-antenna receiver. Let $H_{kj}$ denote the random coefficient of the channel between the $k$-th transmitter and the $j$-th receiver. Suppose $H_{kj}$, $k, j = 1, \ldots, K$, are independent and identically distributed according to $\mathcal{CN}(0, \delta_{kj})$, $k, j = 1, \ldots, K$. Let $p_k$ denote the transmit power for the $k$-th transmitter, where

\[
0 \leq p_k \leq R_k, \quad k = 1, \ldots, K.
\]

Here, $P_k$ represents the power limit for the $k$-th transmitter. Denote $p \triangleq (p_k)_{k=1}^K$. The ergodic rate of the $k$-th pair is given by $r_k(p) = \mathbb{E} \left[ \log \left( 1 + \frac{g_{k,i}^2 p_k}{\sum_{k',j} |H_{kj}|^2 p_k + \sigma_k^2} \right) \right]$, where $\sigma_k$ denotes the variance of the additive complex Gaussian noise at the $k$-th receiver. The ergodic sum-rate of the $K$ pairs is given by $r_0(p) = \sum_{k=1}^K r_k(p)$. The ergodic rate of the $k$-th pair satisfies:

\[
r_k(p) \geq R_k, \quad k = 1, \ldots, K,
\]

where $R_k$ represents the rate requirement for the $k$-th pair. By (18), $r_0(p) \geq \mathbb{E} \left[ \log(1 + \frac{\sum_{k,j} |H_{kj}|^2 p_k}{\sum_{k',j} |H_{kj}|^2 p_k + \sigma_k^2}) \right] \geq r_{0,k}(p_k).$

Thus, a stronger and decoupled version of (19) is given by:

\[
r_{0,k}(p_k) \geq R_k, \quad k = 1, \ldots, K.
\]

We would like to optimize the transmit power $p$ to maximize the ergodic sum-rate $r_0(p)$, subject to the power constraints in (15) as well as the coupled and decoupled individual ergodic rate constraints in (19) and (20), respectively.

Problem 7 (Ergodic Sum-Rate Maximization with Coupled Constraints):

\[
\max_P r_0(p)
\]

s.t. (18), (19).

Problem 8 (Ergodic Sum-Rate Maximization with Decoupled Constraints):

\[
\max_P r_0(p)
\]

s.t. (18), (20).

Problem 7 is one instance of Problem 1. We can choose $\tilde{g}_0$ and $\tilde{g}_{k,i}$, given by (21) and (22), as shown at the top of this page, and obtain a stationary point of Problem 7 using Alg. 1. Problem 8 is one instance of Problem 2. We can choose $\tilde{g}_{k,0}$ and $\tilde{g}_{k,1}$, given by (23) and (24), as shown at the top of this page, and obtain a stationary point of Problem 8 using Alg. 2. Problem 8 has a smaller optimal ergodic sum-rate than Problem 7, but yields a parallel SSCA method with faster convergence speed. Thus, Problem 8 has application when the network topology changes rapidly over time.

V. NUMERICAL RESULTS

In this section, we consider the application examples in Section IV and compare the proposed SSCA and parallel SSCA methods (i.e., Alg. 1 and Alg. 2) with those in [5] through numerical experiments. We set $K = 5$ and $\rho = 0.5$. For simplicity, we choose $P_0 = 100$, $\sigma_k^2 = 1$ and $R_k = 1$ for all $k = 1, \ldots, K$. We choose $\delta_{kj} = 1$ if $k = j$ and $\delta_{kj} = 0.1$ otherwise. We independently generate 50 sample paths of random channel coefficients according to $\mathcal{CN}(0, \delta_{kj})$, $k, j = 1, \ldots, K$, and evaluate the average convergence rates.
\[
\hat{g}_0(p, p^{-1}, H^t) = \sum_{k=1}^{K} \left( \log \left( \sum_{j=1}^{K} |H_{kj}|^2 p_j + \sigma_k \right) - \log \left( \sum_{j=1, j \neq k}^{K} |H_{kj}|^2 p_j + \sigma_k \right) \right)
\]
\[
\hat{g}_k(p, p^{-1}, H^t) = R_k - \log \left( \sum_{j=1}^{K} |H_{kj}|^2 p_j + \sigma_k \right) + \log \left( \sum_{j=1, j \neq k}^{K} |H_{kj}|^2 p_j + \sigma_k \right),
\]
\[
\hat{g}_{k,0}(p_k, p^{t-1}, H^t) = \sum_{m=1}^{K} \left( \log \left( |H_{mk}|^2 p_k + \sum_{j=1, j \neq k}^{K} |H_{mj}|^2 p_j + \sigma_m \right) - \log \left( \sum_{j=1, j \neq m}^{K} |H_{mj}|^2 p_j + \sigma_m \right) \right) \frac{|H_{mk}|^2 (p_k - p^*_k)}{\sum_{l=1, l \neq k}^{K} |H_{lk}|^2 (p_l - p^*_l) + \sigma_m}
\]
\[
\hat{g}_{k,1}(p, p^{t-1}, H^t) = R_k - \log \left( |H_{kk}|^2 p_k + \sum_{j=1, j \neq k}^{K} |H_{kj}|^2 p_j + \sigma_k \right) - \log \left( \sum_{j=1, j \neq k}^{K} |H_{kj}|^2 p_j + \sigma_k \right) - \frac{|H_{kk}|^2 (p_k - p^*_k)}{\sum_{l=1, l \neq k}^{K} |H_{lk}|^2 (p_l - p^*_l) + \sigma_k},
\]

Fig. 1. Convergence rates and computing times.

Table I includes the numbers of iterations and total computing times.

| Problem | Convergence rates | Computing times |
|---------|------------------|-----------------|
| D. P. Bertsekas and J. N. Tsitsiklis, “Gradient convergence in gradient methods with errors,” SIAM Journal on Optimization, vol. 10, no. 3, pp. 627–642, Jan. 2000.
| [2] A. Ruszczynski, “Feasible direction methods for stochastic programming problems,” SIAM Journal on Optimization, vol. 19, no. 1, pp. 220–229, Jan. 1999.
| [3] J. Mairal, “Stochastic majorization-minimization algorithms for large-scale optimization,” June 2013.
| [4] Y. Yang, G. Scutari, D. P. Palomar, and M. Pesavento, “A parallel decomposition method for nonconvex stochastic multi-agent optimization problems,” IEEE Trans. Signal Process., vol. 64, no. 11, pp. 2949–2964, Jun. 2016.
| [5] A. Liu, V. K. N. Lau, and B. Kanarian, “Stochastic successive convex approximation for non-convex constrained stochastic optimization,” IEEE Trans. Signal Process., vol. 67, no. 16, pp. 4189–4203, Aug. 2019.
| [6] O. Mehmeh, K. Huang, B. Gopalakrishnan, A. Konar, and N. D. Sidirooulos, “Feasible point pursuit and successive approximation of non-convex cpqps,” IEEE Signal Process. Lett., vol. 22, no. 7, pp. 804–808, Jul. 2015.
| [7] A. H. Phan, H. D. Tuan, H. H. Kha, and D. T. Ngo, “Nonsmooth optimization for efficient beamforming in cognitive radio multicast transmission,” IEEE Trans. Signal Process., vol. 60, no. 6, pp. 2941–2951, Jun. 2012.

TABLE I. Numbers of iterations and total computing times at \( \|p^* - p\|^2 / \|p^*\|^2 = 0.02 \).

| Algorithm | Number of iterations | Computing time | Computing time | Computing time |
|-----------|---------------------|----------------|----------------|----------------|
| Alg. 1    | 1500                | 1150           | 958            |                |
| SSCA [5]  | 2390                | 123            | 725            |                |
| Alg. 2    | 1084                | 155            | 378            |                |

VI. CONCLUSION

In this letter, we considered the general stochastic optimization problem with expectations in both the objective and constraint functions. We proposed a SSCA method and a parallel SSCA method to obtain stationary points of the general stochastic optimization problem and its special case with decoupled constraint functions, respectively. We provided application examples of the proposed methods and demonstrated the advantages of the proposed methods in terms of convergence rate and computational complexity.