Control
and Cybernetics

VOL. 13 (1984) No. 1–2

A recursive quadratic programming algorithm for constrained stochastic programming problems

by

ANDRZEJ RUSZCZYŃSKI

Institute of Automatic Control
Technical University of Warsaw
ul. Nowowiejska 15/19
00-665 Warszawa, Poland

A stochastic approximation algorithm for constrained problems which corresponds to deterministic recursive quadratic programming methods is defined. The main feature of the algorithm is that random gradient estimates are recursively averaged by an auxiliary filter and the averages obtained are usen in quadratic subproblems which generate successive search directions. Convergence with probability one to the set of Kuhn-Tucker points is proven under typical noise conditions.

1. Introduction

Let F, g_i, $i=1, ..., m_g$, and h_i, $i=1, ..., m_h$, be continuously differentiable functions from R^n to R^1. Consider a constrained optimization problem:

$$\begin{aligned} &\text{minimize } F(x) \\ &\text{subject } to \\ &\qquad g_i(x)\leqslant 0, \quad i=1, ..., m_g, \\ &\qquad h_i(x)=0, \quad i=1, ..., m_h. \end{aligned} \tag{1.1}$$

We assume that the values of the objective function F and its gradient ∇F are observed (computed) in the presence of stochastic noise. A typical problem of this kind is that with $F(x)=E\{f(x, \theta)\}$, where θ is a stochastic parameter. In such a problem one can observe $f(x, \theta)$ for various samples of θ but the average cost $E\{f(x, \theta)\}$ is usually difficult to calculate.

Since it is not possible to compute the exact values of F and ∇F at a given x it is necessary to use for the solution of (1.1) stochastic approximation algorithms (see e.g. [2, 7] and the references therein). Various algorithms of this type have been suggested for constrained problems: the projection method [2, 9, 10], feasible direction methods [1, 11, 17], penalty methods [3, 5, 13, 15], Lagrangian and penalty-multiplier methods [4, 8, 9, 12, 15]. The projection method and feasible direction methods allow for inequality constraints only; the Lagrangian and the penalty-multiplier methods require simultaneous iteration of both primal and dual variables, which causes several difficulties in practice.

The objective of this paper is to present a new stochastic approximation algorithm for (1.1), which corresponds to recursive quadratic programming methods of nonlinear programming (see e.g. [16]). The algorithm extends to constrained problems the concept of the unconstrained stochastic approximation method with averaging studied in [18].

In section 2 we define the algorithm and formulate relevant assumptions. Sections 3 and 4 are devoted to the derivation of some preliminary results and in section 5 we prove the main convergence theorem. We use $|\cdot|$ to denote the Euclidean norm in R^n. We denote by $U_\delta(x)$ the δ-neighborhood of x, i.e. $U_\delta(x)=\{y: |y-x|\leqslant\delta\}$. If $V\subset R^n$ then $U_\delta(V)=\bigcup_{x\in V} U_\delta(x)$. For a closed convex set Z we denote by $\pi_Z(\cdot)$ the orthogonal projection on Z. For a sequence $\{x^k\}_{k=0}^{\infty}$ we use $\mathscr{K}$ to denote an infinite subset of the set of natural numbers and $\{x^k\}_{k\in\mathscr{K}}$ denotes the subsequence associated with κ. We use Ω to denote a probability space and ω to denote elements of Ω. The abbreviation "wp1" is used for "with probability one".

2. The algorithm and the assumptions

Define an auxiliary quadratic programming problem with the decision vector $d\in R^n$ and the parameters $x\in R^n$ and $z\in R^n$:

$$\begin{aligned}&\text{minimize } [\eta(d)=\langle z, d\rangle+\tfrac{1}{2}|d|^2]\\ &\text{subject to}\\ &d\in D(x)=\{d: g_i(x)+\langle\nabla g_i(x), d\rangle\leqslant 0, \quad i=1,\dots,m_g,\\ &\qquad\qquad h_i(x)+\langle\nabla h_i(x), d\rangle=0, \quad i=1,\dots,m_h\}.\end{aligned}\tag{2.1}$$

The solution of (2.1) will be denoted by $d(x, z)$.

Let $\kappa>0$. Define the set

$$X_\kappa=\{x: g_i(x)\leqslant\kappa, \quad i=1,\dots,m_g, \quad |h_i(x)|\leqslant\kappa, \quad i=1,\dots,m_h\}.$$

The feasible set of (1.1) is denoted by X.

Let us consider now the following algorithm for the solution of (1.1):

$$z^{k+1}=z^k+a\tau_k(\xi^k-z^k),\tag{2.2}$$

$$d^k=d(x^k, z^{k+1}),\tag{2.3}$$

$$x^{k+1}=\begin{cases}x^k+\tau_k d^k & \text{if } x^k+\tau_k d^k\in X_\kappa, \quad \text{(a)}\\ x^k & \text{otherwise,} \quad \text{(b)}\end{cases}\tag{2.4}$$

where $z^0\in R^n$, $x^0\in X_\kappa$. The vector ξ^k appearing in (2.2) is a stochastic estimate of the gradient $\nabla F(x^k)$, i.e.

$$\xi^k=\nabla F(x^k)+r^k,$$

where r^k denotes a stochastic noise. The parameter τ_k is a nonnegative step coefficient and a is a positive constant. Algorithm (2.2)—(2.4) will be called the *stochastic recursive quadratic programming method.* Let us note that if the estimate z^{k+1} produced by the auxiliary filter (2.2) is equal to the actual gradient $\nabla F(x^k)$ then the algorithm becomes identical with the deterministic *linearization method* of [16]. It is also worth noting that in the unconstrained case we have $d^k = -z^{k+1}$, i.e. the algorithm reduces to the *method with averaging* analysed in [18].

In [6] another stochastic version of the linearization method was proposed for deterministic nondifferentiable problems. Instead of a τ_k in (2.2) coefficients $\rho_k \to 0$ were used and an additional assumption that $\tau_k/\rho_k \to 0$ was imposed. Hence, for large k the changes in x^n become neglectible, as compared with the operation of the filter (2.2), which results in the increasing accuracy of the approximation of $\nabla F(x^k)$ with z^{k+1}. We show that this is not necessary for convergence; one may have changes in x comparable to those in z. Our technique of convergence analysis, based on a special Liapunov function, leads to the conclusion that convergence occurs provided a is greater than some constant $a_{\min}$, which is unfortunately rather hard to estimate.

We shall take the following assumptions:

(H1) the function F and all functions g_i and h_i are continuously differentiable;

(H2) there exist constants C_0, C_i^g, C_i^h such that for all $x', x'' \in R^n$ we have

$$|\nabla F(x') - \nabla F(x'')| \leqslant C_0 |x' - x''|,$$

$$|\nabla g_i(x') - \nabla g_i(x'')| \leqslant C_i^g |x' - x''|,$$

$$|\nabla h_i(x') - \nabla h_i(x'')| \leqslant C_i^h |x' - x''|;$$

(H3) The set X_κ is bounded and there exists a constant C such that for every $x \in X_\kappa$ and any $d \in D(x)$ one has $|d| \leqslant C$.

(H4) for any $x \in X_\kappa$ and any $d \in D(x)$ the gradients $\nabla h_i(x)$ and these gradients $\nabla g_i(x)$ for which $g_i(x) + \langle \nabla g_i(x), d \rangle = 0$ are linearly independent;

(H5) the set $F(X^*)$, where X^* is the set of Kuhn-Tucker points of (1.1), does not contain any segment of nonzero langth;

(H6) $\tau_k \geqslant 0$ for $k = 0, 1, \ldots$ and $\tau_k \to 0$ wp1;

(H7) $\sum_{k=0}^{\infty} \tau_k = \infty$ wp1;

(H8) there exists $T > 0$ such that

$$\lim_{k \to \infty} \max_{l \in L(k,T)} \left| \sum_{i=k}^{l-1} \tau_i r^i \right| = 0 \text{ wp1},$$

where

$$L(k, T) = \left\{ 1 \geqslant k : \sum_{i=h}^{l-1} \tau_i \leqslant T \right\}.$$

(H9) $a > \frac{1}{2} (C_\delta + \sqrt{(C_\delta)^2 + (C_0)^2}$ where C_δ will be specified in sec. 5.

Let us note that assumptions (H6), (H7) and (H8), concerning the sequence of step coefficients $\{\tau_k\}$ and the noise $\{r^k\}$ are identical with the conditions used in [9] for various stochastic approximation algorithm. In [9] one can find a through discussion of these rather weak, but involved conditions.

Under the above assumptions we shall establish the convergence wp1 of the sequence $\{x^k\}$ to the set of Kuhn-Tucker points of (1.1). In the convergence analysis we shall use the following theorem from [14].

THEOREM 1. *Let $Y^* \subset R^m$. Let $\{y^0\}$ be a bounded sequence in R^m which satisfies the following conditions:*

(a) *if a subsequence $\{y^k\}_{k \in K}$ converges to $y' \in Y^*$, then $|y^{k+1} - y^k| \to 0$ for $k \in \mathscr{K}$;*

(b) *if a subsequence $\{y^k\}_{k \in \mathscr{K}}$ converges to $y' \notin Y^*$ then there exists $\varepsilon_0 > 0$ such that for all $\varepsilon \in (0; \varepsilon_0]$ and all $k \in \mathscr{K}$ the index*

$$s(k, \varepsilon) = \min \{l > k : |y^l - y^k| > \varepsilon\}$$

is finite;

(c) *the exists a continuous function $W(y)$ such that if $\{y^k\}_{k \in K} \to y' \notin Y^*$ then we can find $\varepsilon_1 > 0$ such that for all $\varepsilon \in (0; \varepsilon_1]$ we have*

$$\overline{\lim_{k \in K}}\, W(y^{s(k, \varepsilon)}) < \lim_{k \in \mathscr{K}} W(y^k),$$

where $s(k, \varepsilon)$ is defined as in (b);

(d) *the set $W(Y^*)$ does not contain any segment of nonzero length.*

Then the sequence $\{W(y^k)\}$ converges and all accumulation points of the sequence $\{y^k\}$ belong to X^*.

We shall call the function $W(\cdot)$ *the Liapunov function* and the set Y^* will be called *the solution set*.

In what follows we shall prove that for almost all $\omega \in \Omega$ the paths $\{x^k(\omega), z^k(\omega)\}$ of the sequence $\{x^k, z^k\}$, generated by (2.2)—(2.4), satisfy the assumptions of Theorem 1.

3. Properties of the auxiliary QP subproblems

Let us denote by $\eta(x, z)$ the optimal value of (2.1) and by $\lambda(x, z)$ and $\mu(x, z)$ the multiplier vectors that correspond to the inequality and equality constraints in (2.1).

LEMMA 1. *Assume* (H1) *through* (H4). *Let $x \in X_\kappa$ and $z \in R^n$ be fixed. Then there exist $\bar{\varepsilon} > 0$, C_d, C_λ, C_μ such that for all $(x^1, z^1) \in U_{\bar{\varepsilon}}(x, z)$, $(x^2, z^2) \in U_{\bar{\varepsilon}}(x, z)$ we have*

$$|d^1 - d^2| \leq C_d (|x^1 - x^2| + |z^1 - z^2|),$$
$$|\lambda^1 - \lambda^2| \leq C_\lambda (|x^1 - x^2| + |z^1 - z^2|),$$
$$|\mu^1 - \mu^2| \leq C_\mu (|x^1 - x^2| + |z^1 - z^2|),$$

where $d^j = d(x^j, z^j)$, $\lambda^j = \lambda(x^j, z^j)$, $\mu^j = \mu(x^j, z^j)$ $\quad j = 1, 2$.

P r o o f. The proof follows immediately from [7, thm. 3.1] and will be therefore omitted.

LEMMA 2. *Assume* (H1) *through* (H4). *Let* $x \in X_K$, $z \in R^n$ *be fixed and let* $\bar{\varepsilon}$ *be defined as in Lemma* 1. *Then for any* $\varepsilon \in (0, \bar{\varepsilon})$ *we can find* C *such that if* $(x^1, z^1) \in U_\varepsilon(x, z)$ *and* $(x^2, z^2) \in U_\varepsilon(x, z)$ *then*

$$\eta(x^2, z^2) - \eta(x^1, z^1) \geq \langle d^1, z^2 - z^1 \rangle - \langle z^1 + d^1, x^2 - x^1 \rangle + \\ - |d^1|\,|x^2 - x|^1 \left(\sum_{i=1}^{m_g} \lambda_i^1 C_i^g + \sum_{i=1}^{m_h} \mu_i^1 C_i^h \right) - C_\varepsilon^2, \qquad (3.1)$$

where $d^1 = d(x^1, z^1)$, $\lambda^1 = \lambda(x^1, z^1)$, $\mu^1 \equiv \mu(x^1, z^1)$.

P r o o f. Consider the Lagrange function for (2.1):

$$L(d, \lambda, \mu, x, z) = \langle z, d \rangle + \frac{1}{2} |d|^2 + \sum_{i-1}^{m_g} \lambda_i (g_i(x) + \langle \nabla g_i(x), d) + \\ + \sum_{i=1}^{m_h} \mu_i (h_i(x) + \langle \nabla h_i(x), d \rangle).$$

Let d^1, λ^1, μ^1 and d^2, λ^2, μ^2 be the solutions and Lagrange multipliers for (2.1) at (x^1, z^1) and (x^2, z^2), respectively. Then

$$\eta(x^2, z^2) - \eta(x^1, z^1) = L(d^2, \lambda^2, \mu^2, x^2, z^2) - L(d^1, \lambda^1, \mu^1, x^1, z^1) \geq \\ \geq L(d^2, \lambda^1, \mu^1, x^2, z^2) - L(d^1, \lambda^1, \mu^1, x^1. z^1).$$

By Lemma 1, the latter difference may be easily estimated from below by the expansion of $L(., \lambda^1, \mu^1, .,.,)$ at (d^1, x^1, z^1). Using then the necessary condition of optimality $\nabla_d L = 0$ and the assumption (H2), one immediately obtains the required result.

4. Some properties of the sequences $\{x^k\}$ and $\{z^k\}$.

We shall address at first the question of the feasibility of all cluster points of the sequence $\{x^k\}$.

LEMMA 3. *Assume* (H1), (H2), (H3) *and* (H5). *Then* wp1 *one can find* $k_0 \geq 0$ *such that* $x^k + \tau_k d^k \in X_K$ *for all* $k \geq k_0$.

P r o o f. By construction, $x^k \in X_K$ for $k \geq 0$. According to (H3), the directions d^k are well-defined and $|d^k| \leq C$, $k = 0, 1, \ldots$, where C is a certain constant. It follows from (H2) that

$$g_i(x^k + \tau_k d^k) \leq g_i(x^k) + \tau_k \langle \nabla g_i(x^k), d^k \rangle + C_i^g \tau_k^2 |d^k|^2.$$

Hence, by the definition of d^k, one has

$$g_i(x^k+\tau_k d^k)\leqslant(1-\tau_k)\, g_i(x^k)+C_i^g C^2 \tau_k^2. \tag{4.1}$$

By virtue of (H6), wp1 exists $k_0\geqslant 0$ such that $\tau_k\leqslant\max(1, \kappa/C_i^g C^2)$ for $k\geqslant k_0$, $i=1, \ldots, m_g$. Then

$$g_i(x^k+\tau_k d^k)\leqslant(1-\tau_k)\, g_i(x^k)+\tau_k \kappa\leqslant\kappa,$$

since $g_i(x^k)\leqslant\kappa$ by construction. By a similar argument for equality constraints (expressed as two inequalities) we obtain the assertion of the lemma. ∎

The above lemma shows that wp1 the iterations of the algorithm are performed according to (2.4)—(a), starting from some index k_0 (which may depend on the event ω).

LEMMA 4. *Assume* (H1) *trough* (H4) *and* (H6), (H7), (H8). *Then there is a null set* Ω_0 *such that if* $\omega\notin\Omega_0$ *then all accumulation points of the sequence* $\{x^k(\omega)\}$ *belong to* X.

P r o o f. Let Ω_0 be the null set of (H6), (H7). Let $\omega\notin\Omega_0$ be fixed and let $\{x^k\}$ be the path that corresponds to ω. We shall use Theorem 1, setting the Liapunov function"

$$V(x)=\sum_{i=1}^{m_g}\max(0, g_i(x))+\sum_{i=1}^{m_h}|h_i(x)|,$$

and considering X as the "solution set".

Let us verify the conditions of Theorem 1.

Condition (a). It follows from (H3) and (H6) that $|x^{k+1}-x^k|\leqslant\tau_k|d^k|\to 0$ and thus (a) holds.

Condition (b). Let $\{x^k\}_{k\in\mathscr{K}}\to x'\notin X$. Suppose by contradiction that for any $\varepsilon_0>0$ one can find $\varepsilon\in(0;\varepsilon_0)$ and $k_1\in\mathscr{K}$ such that $|x^i-x^k|\leqslant\varepsilon$ for all $i\geqslant k_1$. Let k_0 be such that for $i\geqslant k_0$ one has $x^i+\tau_i d^i\in X_\kappa$ and $\tau_i\leqslant 1$. The index k_0 exists by Lemma 3. Let $k\geqslant\max(k_0, k_1)$. It follows from (4.1) that for $i\geqslant k$ we have

$$g_j(x^{i+1})\leqslant(1-\tau_i)\, g_j(x^i)+C_1\tau_i^2, \quad j=1, \ldots, m_g,$$

where C_1 does not depend on i, j. Hence

$$\max(0, g_j(x^{i+1}))\leqslant(1-\tau_i)\max(0, g_j(x^i))+C_1\tau_i^2, \quad j=1, \ldots, m_g.$$

By a similar argument for equality constraints we obtain

$$|h_j(x^{i+1})|\leqslant(1-\tau_i)\,|h_j(x^i)|+C_2\tau_i^2, \quad j=1, \ldots, m_h.$$

Thus, for all $i\geqslant k$ the following inequality holds

$$V(x^{i+1})\leqslant V(x^i)+\tau_i[-V(x^i)+C_3\tau_i],$$

where C_3 is a certain constant. Since $|x^i-x^{k_1}|\leqslant\varepsilon$ for $i\geqslant k_1$, then $|x^i-x'|\leqslant 2\varepsilon$ for $i\geqslant k$. Hence

$$V(x^{i+1})\leqslant V(x^i)+\tau_i[-V(x')+C_3\tau_i+C\varepsilon],$$

where C is the Lipschitz constant of V in $U_{2\varepsilon}(x')$. Therefore for any $\varepsilon \in (0; \varepsilon_1]$, any $k \geqslant \max(k_0, k_1)$ and any $l \geqslant k$ one has:

$$V(x^{l+1}) \leqslant V(x^k) + \sum_{i=k}^{l} [-V(x') + C_3 \tau_i + C\varepsilon] \tau_i. \tag{4.2}$$

Observe that if $x' \notin X$ then $V(x') > 0$. Take ε_0 small enough that $C\varepsilon \leqslant V(x')/3$, and let k be so large that $C_3 \tau_i \leqslant V(x')/3$ for $i \geqslant k$. Then it follows from (H7) and (4.2) that $V(x^{l+1}) \to -\infty$ as $l \to \infty$, which contradicts the nonnegativity of $V(\cdot)$. Condition (*b*) must therefore be satisfied.

Condition (c). Let $\{x^k\}_{k \in K} \to x' \notin X$. Let

$$s(k, \varepsilon) = \min\{l > k : |x^l - x^k| > \varepsilon\}.$$

If $k \geqslant k_0$ and $l < s(k, \varepsilon)$ then inequality (4.2) is true. Hence

$$V(x^{s(k, \varepsilon)}) \leqslant V(x^k) + \sum_{i=k}^{s(k, \varepsilon)-1} [-V(x') + C_3 \tau_i + C\varepsilon] \tau_i.$$

Let $V(x') = \delta > 0$. Let us choose $\varepsilon_1 > 0$ and $k \in \mathscr{K}$ such that $-V(x') + C_3 \tau_i + C\varepsilon \leqslant -\delta/2$ for $i \geqslant k$, $\varepsilon \leqslant \varepsilon_1$. Then

$$V(x^{s(k, \varepsilon)}) \leqslant V(x^k) - \frac{\delta}{2} \sum_{i=k}^{s(k, \varepsilon)-1} \tau_i.$$

It follows from the definition of $s(k, \varepsilon)$ that

$$\varepsilon \leqslant |x^{s(k, \varepsilon)} - x^k| \leqslant \sum_{i=k}^{s(k, \varepsilon)-1} \tau_i |d^i| \leqslant C \sum_{i-k}^{s(k, \varepsilon)-1} \tau_i.$$

Combining the two preceding inequalities we obtain

$$V(x^{s(k, \varepsilon)}) \leqslant V(x^k) - \delta\varepsilon/2C,$$

which proves that condition (c) holds.

Condition (d). By definition, $V(X) = \{0\}$ and thus (d) holds.

By Theorem 1, all cluster points of $\{x^k\}$ belong to X, which was set out to prove. ■

Let us derive now a simple but important property of the sequence $\{z^k\}$. Define the sets

$$\nabla F(X_\kappa) = \{\nabla F(x) : x \in X_\kappa\}, \quad Z_\kappa = \mathrm{co}\{\nabla F(X_\kappa)\}.$$

LEMMA 5. *Assume* (H1), (H3) *and* (H6), (H7), (H8). *Then* wp1 *for any* $\delta > 0$ *one can find an index* k_0 *such that* $z^k \in U_\delta(Z_\kappa)$ *for all* $k \geqslant k_0$.

Proof. Let Ω_0 be the null set of (H6)—(H8) and let $\omega \notin \Omega_0$ be fixed. Define the sequences $\{z_1^k\}$ and $\{z_1^k\}$ by

$$z_1^{k+1} = z_1^k + a\tau_k(\nabla F(x^k) - z_1^k), \quad z_1^0 = 0,$$
$$z_2^{k+1} = z_2^k + a\tau_k(r^k - z_2^k), \quad z_2^0 = z^0.$$

Obviously, $z^k = z_1^k + z_2^k$ for all $k \geq 0$. One can easily prove that $\{z_2^k\} \to 0$ a.s., under (H6)—(H8). To this end one can e.g. use Theorem 1 in a way similar to that of Lemma 4, setting $V(z) = |z|^2$ and $Z^* = \{0\}$, or theorems 4.7.1 and 2.3.1 from [9]. Let us consider the sequence $\{z_1^k\}$. Let $\pi_{Z_\kappa}(z_1^k)$ be the projection of z_1^k on Z_κ and let

$$\delta_k = |z_1^k - \pi_{Z_\kappa}(z_1^k)|$$

be the distance from z_1^k to Z_κ. We have

$$z_1^{k+1} = (1 - a\,\tau_k)\, z_1^k + a\,\tau_k\, \nabla F(x^k) = (1 - a\,\tau_k)\, \pi_{Z_\kappa}(z_1^k) + a\,\tau_k\, \nabla F(x^k) + \\ + (1 - a\,\tau_k)\,(z_1^k - \pi_{Z_\kappa}(z_1^k)).$$

Take k_1 such that $a\,\tau_k \leq 1$ for $k \geq k_1$. The vector

$$v^{k+1} = (1 - a\,\tau_k)\, \pi_{Z_\kappa}(z_1^k) + a\,\tau_k\, \nabla F(x^k)$$

is for $k \geq k_1$ a convex combination of elements of Z_κ, and thus belongs to Z_κ. Consequently

$$\delta_{k+1} \leq |z_1^{k+1} - v^{k+1}| = (1 - a\,\tau_k)\, |z_1^k - \pi_{Z_\kappa}(z_1^k)\,| = (1 - a\,\tau_k)\, \sigma_k\,.$$

It follows from the above inequality and (H6), (H7) that $\sigma_k \to 0$. Since $z_2^k \to 0$ then also the distance from z^k to the set Z_κ tends to 0, as $k \to \infty$. The lemma has been proved. ■

5. Convergence analysis

Before proceeding to the convergence analysis we shall specify the constant C_δ in (H9).

Let

$$p(x, z) = \sum_{i=1}^{m_g} \lambda_i(x, z)\, C_i^g + \sum_{i=2}^{m_h} |\mu_i(x, z)\,|\, C_i^h.$$

It follows from Lemma 1 that $p(x, z)$ is continuous on $X_\kappa \times U_\delta(Z_\kappa)$ for any $\delta > 0$. Since, according to (H2) and (H3), the sets X_κ and $U_\delta(Z_\kappa)$ are bounded, then the constant

$$C_\delta = \max_{\substack{x \in X_\kappa \\ z \in U_\delta(Z_\kappa)}} p(x, z)$$

is finite. Therefore we can make the following assumption: (H9) there exists $\delta > 0$ such that $a > \frac{1}{2}\,(C_\delta + \sqrt{(C_\delta)^2 + (C_0)^2})$.

It is worth noting that in the unconstrained problem we have $C_\delta = 0$ and (H9) takes on the form $a > C_0/2$, which is identical with the assumption used in [18] for the unconstrained version of our method. If the constraints are linear then we have also $C_\delta = 0$. The essence of (H9) is that the filter (2.2) should be fast enought to keep up with the varying gradient of the objective function.

Let X^* be the set of Kuhn-Tucker points of (1.1). Define the solution set"

$$Y^* = \{(x, z) \colon x \in X^*, \quad z = \nabla F(x)\}$$

and the "Liapunov function"

$$W(x, z)=a F(x)-\eta(x, z)+\frac{1}{2}|\nabla F(x)-z|^2.$$

We shall prove that wp1 the sequence $\{x^k, z^k\}$ satisfies the assumptions of Theorem 1 and thus converges to Y^*. It can be seen that in the unconstrained case one has $\eta(x, z)=-|z|^2/2$ and thus the function $W(x, z)$ becomes identical with the Liapunov function from [18].

Let Ω_0 be the null set of (H6)—(H8) and let $\omega \notin \Omega_0$ be fixed. Consider the path $\{y^k\}=\{x^k, z^k\}$ generated by (2.2)—(2.4). For any $k \geqslant 0$ and any $\varepsilon>0$ define the set

$$I(k, \varepsilon)=\{\bar{l} \geqslant k : |y^i-y^k| \leqslant \varepsilon \quad \text{for } k \leqslant i \leqslant l\}.$$

LEMMA 6. *Assume* (H1)—(H4) *and* (H6)—(H8). *Let* $\omega \notin \Omega_0$ *be fixed. If a subsequence* $\{y^k\}_{k \in \mathscr{K}} \to y' \notin Y^*$ *then there exist* C, $\gamma>0$, $\varepsilon_m>0$ *and* k_m *such that for any* $k \in \kappa$, $k \geqslant \mathscr{K}_m$, *any* $\varepsilon \in (0; \varepsilon_m]$ *and any* $l \in I(k, \varepsilon)$ *one has*

$$W(y^l)-W(y^k) \leqslant \left(-\gamma+C\varepsilon+C \frac{\left|\sum_{i=k}^{l-1} \tau_i r_i\right|}{\sum_{i=k}^{l-1} \tau_i}\right) \sum_{i=k}^{l-1} \tau_i+C\varepsilon^2. \tag{5.1}$$

P r o o f. Since $\{y^k\}_{k \in \mathscr{K}} \to y'$ then all quantities of the form $|\nabla F(x^k)|$, $|z^k|$, $|d^k|$ are uniformly bounded for $k \in \mathscr{K}$. For simplicity, all constants independent of k and ε will be denoted by C. If $\bar{l} \in I(k, \varepsilon)$ then for $k \leqslant i \leqslant \bar{l}$ the quantities of the form $|x^i-x^k|$, $|z^i-z^k|$, $|\nabla F(x^i)-\nabla F(x^k)|$ can be bounded by $C\varepsilon$.

Let us estimate from above the three parts of the difference $W(y^l)-W(y^k)$.

Part 1. We have

$$F(x^l)-F(x^k) \leqslant \langle \nabla F(x^k), x^l-x^k \rangle+C_0 \varepsilon^2=\sum_{i=\kappa}^{l-1} \tau_i \langle \nabla F(x^k), d^i \rangle+C_0 \varepsilon^2. \tag{5.2}$$

Let $\bar{\varepsilon}$ be the radius of the neighborhood of $y'=(x', z')$ in which the assertions of Lemma 1 hold, and let $\varepsilon_0<\varepsilon/2$. Take k_0 such that $|x^k-x'| \leqslant \varepsilon_0$ and $|z^k-z'| \leqslant \varepsilon_0$ for all $k \geqslant k_0$, $k \in \mathscr{K}$. Then for all $\varepsilon \leqslant \varepsilon_0$, $k \geqslant k_0$, $k \in \mathscr{K}$ and $i \in I(k, \varepsilon)$ we have: $|x^i-x'| \leqslant$ $\leqslant |x^i-x^k|+|x^k-x'| \leqslant 2\varepsilon_0 \leqslant \bar{\varepsilon}$, and similarly $|z^i-z'| \leqslant \bar{\varepsilon}$. Hence, according to Lemma 1, one has $|d^i-d^k| \leqslant C\varepsilon$. Therefore (5.2) implies that

$$F(x^l)-F(x^k) \leqslant (\langle \nabla F(x^k), d^k \rangle+C\varepsilon) \sum_{i=k}^{l-1} \tau_i+C_0 \varepsilon^2. \tag{5.3}$$

Part 2. It follows from Lemma 2, that

$$-\eta(x^l, z^l)+\eta(x^k, z^k) \leqslant -\langle d^k, z^l-z^k \rangle+\langle z^k+d^k, x^l-x^k \rangle+$$
$$+|d^k|\,|x^l-x^k| \left(\sum_{j=1}^{m_g} \lambda_j^k C_j^g+\sum_{j=1}^{m_h} |\mu_j^k C_j^h|\right)+C\varepsilon^2. \tag{5.4}$$

By virtue of Lemma 5, there is k_0 such that $z^k \in U_\delta(Z_\kappa)$ for all $k \geqslant k_0$. Then

$$\sum_{j=1}^{m_g} \lambda_j^k C_j^g + \sum_{j=1}^{m_h} |\mu_j^k| C_j^h \leqslant C_\delta$$

for $k \geqslant k_0$. Therefore from (5.4) we obtain

$$-\eta(x^l, z^l) + \eta(x^k, z^k) \leqslant -\langle d^k, z^l - z^k\rangle + \langle z^k + d^k, x^l - x^k\rangle + \\ + C_\delta |d^k| |x^l - x^k| + C\varepsilon^2. \qquad (5.5)$$

Let us note that for $l \in I(k, \varepsilon)$ the difference $z^l - z^k$ may be expressed as

$$z^l - z^k = a(\nabla F(x^k) - z^k) \sum_{i=k}^{l-1} \tau_i + a \sum_{i=k}^{l-1} \tau_i r^i + \\ + a \sum_{i=k}^{l-1} \tau_i (\nabla F(x^i) - \nabla F(x^k) + z^k - z^i), \qquad (5.6)$$

where the last term can be bounded by $C\varepsilon \sum_{i=k}^{l-1} \tau_i$. Similarly,

$$x^l - x^k = d^k \sum_{i=k}^{l-1} \tau_i + \sum_{l=k}^{l-1} (d^i + d^k), \qquad (5.7)$$

with the second term bounded by $C\varepsilon \sum_{i=k}^{l-1} \tau_i$. Expressions (5.6) and (5.7) when applied to (5.5) give

$$-\eta(x^l, z^l) + \eta(x^k, z^k) \leqslant (-a\langle d^k, \nabla F(x^k) - z^k\rangle + \langle z^k + d^k, d^k\rangle + \\ + C_\delta |d^k|^2 + C\varepsilon) \sum_{i=k}^{l-1} \tau_i + C \left| \sum_{i=k}^{l-1} \tau_i r^i \right| + C\varepsilon^2. \qquad (5.8)$$

Part 3. We have

$$\frac{1}{2} |\nabla F(x^l) - z^l|^2 - \frac{1}{2} |\nabla F(x^k) - z^k|^2 = \langle \nabla F(x^k) - z^k, \nabla F(x^l) - \nabla F(x^k)\rangle - \\ - \langle \nabla F(x^k) - z^k, z^l - z^k\rangle + \frac{1}{2} |\nabla F(x^l) - z^l - \nabla F(x^k) + z^k|^2 \leqslant \\ \leqslant C_0 |\nabla F(x^k) - z| |x^l - x^k| - \langle \nabla F(x^k) - z^k, z^l - z^k\rangle + C\varepsilon^2.$$

After substituting (5.6) and (5.7) for $x^l - x^k$ and $z_l - z^k$ in the above inequality we obtain

$$\frac{1}{2} |\nabla F(x^l) - z^l| - \frac{1}{2} |\nabla F(x^k) - z^k| \leqslant (C_0 |\nabla F(x^k) - z^k| |d^k| + \\ - a |\nabla F(x^k) - z^k|^2 + C\varepsilon) \sum_{i-k}^{l-1} \tau_i + C \left| \sum_{i=k}^{l-k} \tau^i r^i \right| + C\varepsilon^2. \qquad (5.9)$$

Part 4. Let us add (5.3) multiplied by the constant a to (5.8) and (5.9). We obtain

$$W(y^l)-W(y^k)\leqslant[(a+1)\langle z^k, d^k\rangle+(C_\delta+1)\,|d^k|^2+ \\ +C_0\,|\nabla F(x^k)-z^k|\,|d^k|-a\,|\nabla F(x^k)-z^k|^2+C\,\varepsilon]\sum_{i=k}^{l-1}\tau_i+ \\ +C\left|\sum_{i=k}^{l-1}\tau_i\,r^i\right|+C\,\varepsilon^2.$$

For a sufficiently small ε and a sufficiently large k (see the motivation of (5.3)) the following bounds hold:

$$\langle z^k, d^k\rangle\leqslant\langle z', d'\rangle+C\,\varepsilon,\ |d^k|^2\leqslant|d'|^2+C\,\varepsilon,\ |\nabla F(x^k)-z^k|\,|d^k|\leqslant \\ \leqslant|\nabla F(x')-z'|\,|d'|+C\,\varepsilon,-a\,|\nabla F(x^k)-z^k|^2\leqslant-a|\,\nabla F(x')-z'|^2+C\,\varepsilon.$$

Hence

$$W(y^l)-W(y^k)\leqslant[(a+1)\langle z', d'\rangle+(C_\delta+1)\,|d'|^2+C_0\,|\nabla F(x')- \\ -z'|\,|d'|-a\,|\nabla F(x')-z'|^2+C\,\varepsilon]\sum_{i=k}^{l-1}\tau_i+C\left|\sum_{i=k}^{l-1}\tau_i\,r^i\right|+C\,\varepsilon^2. \tag{5.10}$$

By Lemma 4, $x'\in X$. Thus $\alpha d'\in D(x')$ and $\eta(\alpha d')\geqslant\eta(d')$ for $0\leqslant\alpha\leqslant1$. Hence $\langle z'+d', d'\rangle\leqslant0$, i.e.

$$(a+1)\langle z', d'\rangle\leqslant-(a+1)\,|d'|^2.$$

Consequently, we can rewrite (5.10) as

$$W(y^l)-W(y^k)\leqslant[(C_\delta-a)\,|d'|^2+C_0\,|\nabla F(x')-z'|\,|d'|- \\ -a|\nabla F(x')-z'|^2+C\,\varepsilon]\sum_{i=\kappa}^{l-1}\tau_i+C\left|\sum_{i=k}^{l-1}\tau_i\,r^i\right|+C\,\varepsilon^2. \tag{5.11}$$

Consider now the quadratic form

$$\psi(u, v)=(C_\delta-a)\,u^2+C_0\,uv-av^2.$$

It may be easily verified that $\psi(.,.)$ is negatively defined, if (H9) holds. Furthermore, the assumption that $(x', z')\notin Y^*$ implies that $|d'|+|\nabla F(x')-z'|>0$. Thus

$$\psi(|d'|,|\nabla F(x')-z'|)<0.$$

Replacing $\psi(|d'|,|\nabla F(x')-z'|)$ in (5.11) by $-\gamma$, where $\gamma>0$, we obtain the required inequality (5.1). The lemma has been proved., ■

Now we can state the main theorem.

THEOREM 2. *Assume* (H1) *to* (H9). *Then there is a null set* Ω_0 *such that* $\omega\notin\Omega_0$ *implies that*:

1° *all accumulation points of the sequence* $\{x^k(\omega)\}$ *are included in* X^*;

2° $z^k(\omega) - \nabla F(x^k(\omega)) \to 0$, *as* $k \to \infty$;
3° *the sequence* $\{F(x^k(\omega))\}$ *is convergent.*

P r o o f. Let us note that inequality (5.1) derived in Lemma 6. is of the same form as inequality (A.1) from [18], where the unconstrained version of the method was analysed. So one can use Theorem 1 in an identical fashion as in [18] to derive all the three assertions of our theorem. ∎

COROLLARY. *For any convergent subsequence* $\{x^k\}_{k\in\mathcal{K}} \to x^* \in X^*$ *one has*

$$\lim_{k\in K} \lambda(x^k, z^k) = \lambda^*,$$

$$\lim_{k\in K} \mu(x^k, z^k) = \mu^*,$$

where $\lambda(x^k, z^k)$, $\mu(x^k, z^k)$ *are the multipliers in auxiliary* QP *subproblems and* λ^*, μ^* *are the values of multipliers at* x^*.

The above corollary follows immediately from assertions 1 and 2 of theorem 2 and from the stability of QP subproblems (Lemma 1).

6. Conclusions

A stochastic approximation algorithm for constrained problems, which corresponds to deterministic recursive quadratic programming methods, has been presented and the convergence of the algorithm has been proved. The algorithm allows for both inequality and equality constraints and is in fact an extension of the stochastic conjugate gradient method to constrained problems. The assumptions imposed on noise $\{r^k\}$ and gains $\{\tau_k\}$ are typical of the theory of stochastic approximation algorithms. The most restrictive assumptions are (H4), which is necessary for the stability of QP subproblems, and (H9), which limits from below the filter gains by an unknown constant. An interesting problem that calls for explanation is the possibility of replacing (2.2) by the formula $z^{k+1} = z^k + \rho_k(\xi^k - z^k)$ with a new gain sequence $\{\rho_k\}$ such that $\liminf \rho_k/\tau_k \geq a$; this could increase flexibility of the algorithm.

References

[1] BAZHENOV L. G., GUPAL A. M. On a stochastic analog of the feasible direction method. *Kibernetika* 1973, 91–95 [in Russian].

[2] ERMOLIEV YU. M. On the stochastic gradient method and stochastic quasi Fejer sequences. *Kibernetika* 1969, 73–83. [in Russian].

[3] ERMOLIEV YU. M. On a general problem of stochastic programming. *Kibernetika* 1971, 47–50 [in Russian].

[4] Ermoliev Yu. M. Stochastic Programming Methods. Moskow. Nauka, 1976 [in Russian].

[5] Fabian V. Stochastic approximation of constrained minima, Transactions of the 4 th Prague Conference on Information Theory, Statistical Decision Functions and Random Processes 1965. Prague 1967, 277–290.

[6] Gupal A. M. Method for minimizing Lipschitz continuous functions. *Kibernetika* 1980 91–94 [in Russian].

[7] Hager W. W. Lipschitz continuity for constrained processes. *SIAM Journal on Control and Optimization* **17** (1979), 321–338.

[8] Katkovnik V. Ya. Linear Estimates and Stochastic Optimization Problems. Moskow, Nauka, 1976 [in Russian].

[9] Kushner H. J., Clark D. S. Stochastic Approximation Methods for Constrained and Unconstrained Systems. New York, Springer, 1978.

[10] Kushner H. J., Clark D. S. A stochastic approximation projection algorithm for constrained Monte Carlo optimization. A. Niemi (ed.), A Link Between Science and Applications of Automatic Control. Proceedings of the 7 th World Congress of the IFAC. Pergamon Press, 1979, 2425–2433.

[11] Kushner H. J., Gavin T. Stochastic approximation type methods for constrained systems algorithms and numerical results. *IEEE Transactions on Automatic Control*, **AC-119** (1974), 349–357.

[12] Kushner H. J., Kelmanson M. L. Stochastic approximation algorithms of the multiplier type for the sequential Monte Carlo optimization of constrained systems. *SIAM Journal on Control and Optimization*, **14** (1976), 827–842.

[13] Kushner H. J., Sanvicente E. Penalty function methods for constrained stochastic approximation. *Journal of Mathematical Analysis and Applications*, **46** (1974), 499–512.

[14] Nurmiński E. A. Convergence conditions for nonlinear programming algorithms. *Kibernetika* 1972, 79–81 [in Russian].

[15] Poljak B. T, Nonlinear programming methods in the presence of noise. *Mathematical Programming*, **14** (1978), 87–97.

[16] Pshenichny B. N., Danilin Yu. M. Numerical Methods in Extremal Problems. Moskow, Mir Publishers, 1978.

[17] Ruszczyński A. Feasible direction methods for stochastic programming problems, *Mathematical Programming*, **19** (1980), 220–229.

[18] Ruszczyński A., Syski W. Stochastic approximation algorithm with gradient averaging for unconstrained problems. *IEEE Transactions on Automatic Control*, [in press].

Received, March 1983.

Algorytm rekurencyjnego programowania kwadratowego dla zadań programowania stochastycznego z ograniczeniami

W pracy sformułowano algorytm aproksymacji stochastycznej dla zadań optymalizacji z ograniczeniami, który odpowiada deterministycznym metodom rekursywnego programowania kwadratowego. Uśrednianie rekurencyjnie w pomocniczym filtrze losowe estymaty gradientów są wykorzystywane w zadaniach programowania kwadratowego do generacji kolejnych kierunków. Wykazano zbieżność z prawdopodobieństwem 1 do zbioru punktów Kuhna-Tuckera przy typowych założeniach o szumie.

Метод рекурентного квадратического программирования для условных задач стохастического программирования

В работе сформулировано метод стохастической аппроксимации для задач с ограничениями, соответствующий детерминистическим методом рекурентного квадратического программирования. Рекурентно усредняемые вспомогательным фильтром случайные оценки градиентов используются в квадратических подпроблемах для получения последовательных направлений спуска. Доказано сходимость с вероятностью единица к множеству точек Куна-Такера при обычных условиях шума.