International Journal of Pure and Applied Mathematics
Volume 10 No. 2 2004, 169-180

TWO PVT ALGORITHMS OF NONMONOTONE-TYPE

Li-Ping Pang[1], Zun-Quan Xia[2 §], Yu-Lin Dong[3]
[1,2,3]Centre for Optimization Research and Applications (CORA)
Department of Applied Mathematics
Dalian University of Technology
Dalian 116024, Liaoning, P.R. CHINA
[1]e-mail: lipingpang@163.com
[2]e-mail: zqxiazhh@dlut.edu.cn
[3]e-mail: dyl-1@163.com

Abstract: Two PVT (parallel variable transformation) algorithms of nonmonotone type with the nonmonotone synchro-parallel step are presented in this paper. The original PVT algorithm, due to Fukushima (1998), [4], designed for solving unconstrained minimization problems is monotone one. The convergence theorems for the two algorithms are given.

AMS Subject Classification: 90C30, 49M37, 65Y05
Key Words: parallel algorithm, unconstrained minimization, nonlinear programming, convergence, monotonicity of algorithms

1. Introduction

We consider solving unconstrained minimization problems

$$\text{(P)} \qquad \min_{x \in R^n} f(x), \tag{1.1}$$

where $f : R^n \to R$ is a continuously differentiable function. It is known for the most of synchronization parallel algorithms, one loop comprises consists of two iterate steps: parallelization and synchronization. The synchronization step

Received: October 29, 2003
[§]Correspondence author

is designed for guaranteeing a sufficient decrement of the objective function, while the parallel step produces candidate points by simultaneously solving certain subproblems (P_l) for $l = 1, \cdots, p$. Each (P_l) is defined on a subspace of dimension smaller than n, such that p subspaces span the whole space R^n. See for in stance the parallel gradient distribution (PGD) algorithm proposed by Mangasarian [6], the parallel variable distribution (PVD) algorithm proposed by Ferris and Mangasarian [3] that was further studied by Solodov [10], and a general framework, called the parallel variable transformation (PVT) algorithm due to Fukushima, also belongs to this kind of algorithms. For the unconstrained problem, solutions of the subproblems are calculated under a certain sufficient descent condition. Solodov proposed a modified algorithm of PVD type that is a nonmonotone synchronization scheme [10]. In is paper, we will present two nonmonotone schemes of PVT type algorithm that are more suitable for implementation.

This paper is organized as follows. In Section 2, we present a PVT algorithm with the nonmonotone parallel step and the global convergence is given. In Section 3, a nonmonotone PVT algorithm is proposed, and $\theta-$descent condition and global convergence of the algorithm are given.

2. A PVT Algorithm with the Nonmonotone Parallel Step

We first list the following basic notations.

Basic Notations:

k	The number of iterations;
p	The number of parallel processors;
P	$= \{1, 2, \cdots, p\}$;
m_l	A positive integer such that $m_1 + m_2 + \cdots + m_p \geq n$;
$B^{(k)}$	An $n \times (p+1)$ matrix whose columns consisting of $x^{(k)}$ and $A_l^{(k)} y_l + x^{(k)}$, $\quad A_l^{(k)} \in R^{n \times m_l}, \quad l = 1, \cdots, p$;
$z^{(k)}$	$= (z_0^{(k)}, z_1^{(k)}, \cdots, z_p^{(k)})^T \in R^{p+1}$;
$\varphi_l^{(k)}(y_l)$	$\equiv f(A_l^{(k)} y_l + x^{(k)})$;
$\psi^{(k)}(z)$	$\equiv f(B^{(k)} z)$;
$A^{(k)}$	$= [A_1^{(k)}, \cdots, A_p^{(k)}]$.

The original PVT algorithm due to Fukushima, [4], is stated below.

Algorithm I. The parallel variable transformation (PVT) algorithm, see [4].

Step 0. (Initialization) An initial point $x^{(0)} \in R^n$ is given and set $k = 0$.

Step 1. (Parallelization) Choose $A_l^{(k)} \in R^{n \times m_l}$, $l \in P$. Compute

$$y_l^{(k)} \in \operatorname{Arg} \min_{y_l \in R^{m_l}} \varphi_l^{(k)}(y_l), \qquad l \in P. \tag{2.2}$$

If $\nabla \varphi_l^{(k)}(0) = 0$, $\forall l \in P$, then stop, and $x^{(k)}$ is a solution. Otherwise, go to Step 2.

Step 2. (Synchronization) Compute

$$z^{(k)} \in \operatorname{Arg} \min_{z \in R^{p+1}} \psi^{(k)}(z). \tag{2.3}$$

Set $x^{(k+1)} = B^{(k)} z^{(k)}$, $k = k + 1$. Loop at Step 1.

End of Algorithm I.

In this algorithm, it is not necessary to solve accurately (2.2) in Step 1 and (2.3) in Step 2. In fact, it suffices to compute, for each k, vectors $y_l^{(k)}$ satisfying

$$\varphi_l^{(k)}(y_l^{(k)}) \le \varphi_l^{(k)}(0) - \sigma_l^{(k)}(\|\nabla \varphi_l^{(k)}(0)\|), \tag{2.4}$$

where $\sigma_l^{(k)} : R_+^n \to R_+^1$ is a forcing function associated with the function $\varphi_l^{(k)}$, see for instance, [4], and $z^{(k)}$ satisfying

$$\psi^{(k)}(z^{(k)}) \le \min_{1 \le l \le p} \varphi_l^{(k)}(y_l^{(k)}). \tag{2.5}$$

Global convergence of the original PVT algorithm is ensured under some assumptions. We will present an algorithm below in which the parallelization step needs not be monotone and can be combined with nonmonotone schemes similar to [7].

Algorithm II. A PVT algorithm with a nonmonotone parallelization step.

Step 0. (Initialization) It is similar to Algorithm I.

Step 1. (Parallelization) Choose $A_l^{(k)} \in R^{n \times m_l}$, $l \in P$. Compute

$$\varphi_l^{(k)}(y_l^{(k)}) \le \varphi_l^{(k)}(0) - \sigma_l^{(k)}(\|\nabla \varphi_l^{(k)}(0)\|) + \theta_l^{(k)}, \tag{2.6}$$

where $\sigma_l^{(k)} : R_+^n \to R_+^1$ is a forcing function. If $\nabla \varphi_l^{(k)}(0) = 0$, for all $l \in P$, then stop. Otherwise, go to Step 2.

Step 2. (Synchronization) Choose $x^{(k+1)}$ satisfying

$$f(x^{(k+1)}) \leq \min_{1\leq l\leq p} \varphi_l^{(k)}(y_l^{(k)}). \tag{2.7}$$

Loop at Step 1.

End of Algorithm II.

In the PVD algorithm with inexact subproblem solutions, due to Solodov [10], it was pointed out that (2.6) is satisfied when $A_l^{(k)}$ is taken as

$$A_l^{(k)} = \begin{pmatrix} I_l & 0 \\ 0 & D_l^{(k)} \end{pmatrix},$$

where I_l is an $n_l \times n_l$ identity matrix and $D_l^{(k)}$ is an $n_l \times (p-1)$ block diagonal matrix. In the algorithm every candidate solution of the subproblem belongs to an $\varepsilon-$stationary set, [11].

The following assumptions are used in proving the convergence of Algprithm II:

(A1) There exists a constant $\beta > 0$ independent of k such that

$$\|A^{(k)T}x\| \geq \sqrt{\beta}\|x\|, \qquad \text{for all} \quad x \in R^n. \tag{2.8}$$

(A2) $\sum_{k=0}^{\infty} \theta_l^{(k)} < +\infty$, for all $l \in \{1, \cdots, p\}$, where $\theta_l^{(k)} > 0$.

(A3) The Uniformly Forcing Property. For each l, the sequence $\{\sigma_l^{(k)} | k = 0, 1, 2, \dots\}$ of forcing functions in (2.6) satisfies the property that, for any $\varepsilon > 0$, there exists a $\delta > 0$ independent of k such that for all k, $\sigma_l^{(k)}(t) < \delta \Rightarrow t < \varepsilon$. see [4].

Remarks. *(1)* The assumption (A1) was used in [4].

(2) The assumption (A1) is equivalent to saying that the sequence $\{A^{(k)}A^{(k)T}\}_{k=1}^{\infty}$ of $n\times n$ matrices is uniformly positive definite, i. e., there exists a constant $\beta > 0$ independent of k such that for all $x \in R^n$, one has $x^T A^{(k)T} A^{(k)} x \geq \beta\|x\|^2$.

(3) The assumption (A3) was used in [4].

(4) Under the condition (A3), the function $\sigma_l : R_+^n \to R_+^1$ defined by

$$\sigma_l(t) = \inf_k \sigma_l^{(k)}(t),$$

also satisfies the property of a forcing function.

Lemma 2.1. (see [2]) *Let* $\{a^{(k)}\}$ *and* $\{\varepsilon^{(k)}\}$ *be two sequences of nonnegative real numbers with* $\varepsilon^{(k)} \geq 0$, $\sum_{k=0}^{\infty} \varepsilon^{(k)} < \infty$, *and* $0 \leq a^{(k+1)} \leq a^{(k)} + \varepsilon^{(k)}$ *for* $k = 0, 1, \cdots$. *Then the sequence* $\{a^{(k)}\}$ *converges.*

We now state and prove a convergence theorem.

Theorem 2.1. *Suppose the following conditions are satisfied:*

(a) $f \in C^1(R^n)$;

(b) $\inf_{x \in R^n} f(x) = \overline{f} > -\infty$;

(c) Conditions (A1)-(A3) are satisfied.

Then Algorithm II either terminates at a stationary point of (P) or generates an infinite sequence $\{x^{(k)}\}$ *such that* $\lim_{k \to \infty} \nabla f(x^{(k)}) = 0$, *and every cluster of* $\{x^{(k)}\}$ *is a stationary point of (P).*

Proof. Suppose the algorithm terminates, say, at iteration k. Then we have $\nabla \varphi_l^{(k)}(0) = 0$, for all l, which implies $A_l^{(k)T} \nabla f(x^{(k)}) = 0$, $l = 1, \cdots, p$. Since $A^{(k)} = [A_1^{(k)}, \cdots, A_p^{(k)}]$, one has

$$\|A^{(k)T} \nabla f(x^{(k)})\|^2 = \sum_{l=1}^{p} \|A_l^{(k)T} \nabla f(x^{(k)})\|^2 = 0.$$

It then follows from the condition (A1) that $\nabla f(x^{(k)}) = 0$. Thus $x^{(k)}$ is a stationary point of (P).

Suppose the algorithm does not terminate in finite steps and generates an infinite sequence $\{x^{(k)}\}$. By virtue of (2.6), (2.7) and (A3) we have

$$\begin{aligned} f(x^{(k+1)}) - f(x^{(k)}) &\leq \varphi_l^{(k)}(y_l^{(k)}) - \varphi_l^{(k)}(0) \\ &\leq -\sigma_l^{(k)}(\|\nabla \varphi_l^{(k)}(0)\|) + \theta_l^{(k)} \leq -\sigma_l(\|\nabla \varphi_l^{(k)}(0)\|) + \theta_l^{(k)}. \end{aligned} \tag{2.9}$$

Without loss of generality, we let $\inf_{x \in R^n} f(x) = \overline{f} > 0$, since $\inf_{x \in R^n} f(x) = \overline{f} > -\infty$. According to (2.9), one has

$$\begin{aligned} \overline{f} - f(x^{(0)}) \leq f(x^{(k)}) - f(x^{(0)}) &= \sum_{i=0}^{k-1} (f(x^{(i+1)}) - f(x^{(i)})) \\ &\leq -\sum_{i=0}^{k-1} \sigma_l(\|\nabla \varphi_l^{(i)}(0)\|) + \sum_{i=0}^{k-1} \theta_l^{(i)}. \end{aligned}$$

Taking $k \to \infty$, we obtain

$$\overline{f} - f(x^{(0)}) \leq -\sum_{i=0}^{\infty} \sigma_l(\|\nabla \varphi_l^{(i)}(0)\|) + \sum_{i=0}^{\infty} \theta_l^{(i)}.$$

We have by (A2) that $\sum_{i=0}^{\infty} \sigma_l(\|\nabla\varphi_l^{(i)}(0)\|) < +\infty$. This leads to

$$\lim_{k\to\infty} \sigma_l(\|\nabla\varphi_l^{(k)}(0)\|) = 0,$$

for all $l = 1, \cdots, p$, and $\lim_{k\to\infty} \|\nabla\varphi_l^{(k)}(0)\| = 0$, for all $l = 1, \cdots, p$. Therefore, one has

$$\begin{aligned}\lim_{k\to\infty} \|A^{(k)T}\nabla f(x^{(k)})\|^2 &= \lim_{k\to\infty} \sum_{l=1}^{p} \|A_l^{(k)T}\nabla f(x^{(k)})\|^2 \\ &= \lim_{k\to\infty} \sum_{l=1}^{p} \|\nabla\varphi_l^{(k)}(0)\| = 0. \quad (2.10)\end{aligned}$$

According to (A1), we have

$$\|A^{(k)T}\nabla f(x^{(k)})\| \geq \sqrt{\beta}\|\nabla f(x^{(k)})\|. \quad (2.11)$$

Combining (2.10) and (2.11) yields $\lim_{k\to\infty} \nabla f(x^{(k)}) = 0$. In consequence, we have from (2.9) that

$$f(x^{(k+1)}) \leq f(x^{(k)}) - \sigma_l(\|\nabla\varphi_l^{(k)}(0)\|) + \theta_l^{(k)} \leq f(x^{(k)}) + \theta_l^{(k)}.$$

It follows from Lemma 2.1 and (A2) that the sequence $\{f(x^{(k)})\}_{k=1}^{\infty}$ converges. Thus, for each accumulation $\overline{x}$ of the sequence $\{x^{(k)}\}_{k=1}^{\infty}$, we have $\nabla f(\overline{x}) = 0$. The demonstration is completed. □

3. A Nonmonotone PVT Algorithm

We present a nonmonotone PVT algorithm, named Algorithm III, whose parallelization and synchronization steps also need not be monotone.

3.1. Algorithm

Algorithm III. The nonmonotone PVT algorithm.

Step 0. (Initialization) It is similar to Algorithm I.

Step 1. (Parallelization) For each $l \in P$, choose a matrix $A_l^{(k)} \in R^{n\times m_l}$ satisfying $\nabla f(x^{(k)}) \in \text{span}A_l^{(k)}$, and find an approximate solution $y_l^{(k)} \in R^{m_l}$

to (2.2). If there exists at least one $l_0 \in P$ such that $\nabla\varphi_{l_0}^{(k)}(0) = 0$, then stop. If all $l \in P$ such that the condition

$$\varphi_l^{(k)}(y_l^{(k)}) \le \varphi_l^{(k)}(0) - \sigma_l \|\nabla\varphi_l^{(k)}(0)\|^2 + \theta_l^{(k)}, \tag{3.12}$$

is satisfied, where σ_l is a constant, then go to Step 2.

Step 2. (Synchronization) Choose $x^{(k+1)}$ satisfying

$$f(x^{(k+1)}) \le \max_{1\le l\le p} \varphi_l^{(k)}(y_l^{(k)}) + \lambda\|\nabla f(x^{(k)})\|^2, \tag{3.13}$$

where $\lambda > 0$ is a constant. Let $k = k+1$. Loop at Step 1.

End of Algorithm III.

The condition (3.12) is designed for finding an approximate solution of the subproblems in Step 1, required in the parallel step of Algorithm III. More details are given in the next subsection.

3.2. A $\theta-$Descent Direction for Solving Subproblems

At iteration k, it suffices to find a $y_l^{(k)}$ such that (3.12) is satisfied when we minimize each auxiliary function $\varphi_l^{(k)}$ with respect to y_l, starting with the origin, i. e., $y_l = 0$. Suppose $y_l^{(k)}$ is computed by

$$y_l^{(k)} = \alpha_l^{(k)} d_l^{(k)}. \tag{3.14}$$

A direction $d_l^{(k)}$ is said to be $\theta - descent$ if the following gradient-relatedness condition of the direction $A_l^{(k)} d_l^{(k)}$ in the sense of Ortega and Rheinboldt, [8],

$$\nabla\varphi_l^{(k)T}(0) d_l^{(k)} \le -\mu_0 \|\nabla\varphi_l^{(k)}(0)\| \cdot \|d_l^{(k)}\| < 0\,, \tag{3.15}$$

is satisfied, with choices of $\mu_0 > 0$ and $\alpha_l^{(k)} > 0$ obeying a nonmonotone Armijo rule

$$\varphi_l^{(k)}(y_l^{(k)}) - \varphi_l^{(k)}(0) \le \mu_1 \alpha_l^{(k)} \nabla\varphi_l^{(k)T}(0) d_l^{(k)} + \theta^{(k)}\,, \tag{3.16}$$

where $\mu_1 \in (0,\,1)$ is a constant, $\alpha_l^{(k)}$ is the maximum positive number which satisfies (3.16).

Lemma 3.1. *Suppose the following conditions are satisfied:*

(a) *$f \in LC_\rho^1(R^n)$, where $LC_\rho^1(R^n)$ denotes the class of all functions that first partial derivatives are Lipschitz continuous with constant $\rho > 0$;*

(b) $y_l^{(k)}$ is determined by (3.14) with $d_l^{(k)}$ and $\alpha_l^{(k)}$ satisfying (3.15) and (3.16);

(c) $\|A_l^{(k)}\| \le \delta_l$, for all l and k.

Then for each l one has that (3.12) holds.

Proof. From (3.16), if $\widetilde{\alpha}_l^{(k)} = 2\alpha_l^{(k)}$, then the line search must be failing, i. e.,

$$\varphi_l^{(k)}(\widetilde{y}_l^{(k)}) - \varphi_l^{(k)}(0) > (1-\widetilde{\mu}_1)\widetilde{\alpha}_l^{(k)}\nabla\varphi_l^{(k)T}(0)d_l^{(k)} + \theta^{(k)}, \tag{3.17}$$

where $\widetilde{y}_l^{(k)} = \widetilde{\alpha}_l^{(k)} d_l^{(k)}$, $\mu_1 = 1-\widetilde{\mu}_1$. According to the Mean-Value Theorem, one has

$$f(A_l^{(k)}\widetilde{y}_l^{(k)} + x^{(k)}) - f(x^{(k)}) = [A_l^{(k)T}\nabla f(\tau A_l^{(k)}\widetilde{y}_l^{(k)} + x^{(k)})]^T \widetilde{y}_l^{(k)}, \tag{3.18}$$

for some $\tau \in (0,1)$. It follows from (3.17)-(3.18) that

$$\begin{aligned}[A_l^{(k)T}\nabla f(\tau A_l^{(k)}\widetilde{y}_l^{(k)} + x^{(k)})]^T \widetilde{y}_l^{(k)} &> (1-\widetilde{\mu}_1)\nabla\varphi_l^{(k)T}(0)\widetilde{y}_l^{(k)} + \theta^{(k)}\\ &> (1-\widetilde{\mu}_1)\nabla\varphi_l^{(k)T}(0)\widetilde{y}_l^{(k)}.\end{aligned}$$

Thus, we obtain

$$\begin{aligned}&[A_l^{(k)T}\nabla f(\tau A_l^{(k)}\widetilde{y}_l^{(k)} + x^{(k)}) - A_l^{(k)T}\nabla f(x^{(k)})]^T d_l^{(k)}\\ &\quad > (1-\widetilde{\mu}_1)\nabla\varphi_l^{(k)T}(0)d_l^{(k)} - \nabla\varphi_l^{(k)T}(0)d_l^{(k)} \ge -\widetilde{\mu}_1\nabla\varphi_l^{(k)T}(0)d_l^{(k)}.\end{aligned} \tag{3.19}$$

Since ∇f is Lipschitzian with constant ρ, one has

$$\begin{aligned}&[\nabla f(x^{(k)} + \tau A_l^{(k)}\widetilde{y}_l^{(k)}) - \nabla f(x^{(k)})]^T A_l^{(k)} d_l^{(k)}\\ &\quad \le \tau\rho\widetilde{\alpha}_l^{(k)}\|A_l^{(k)}\|^2 \cdot \|d_l^{(k)}\|^2 \le \tau\rho\delta_l^2\widetilde{\alpha}_l^{(k)}\|d_l^{(k)}\|^2.\end{aligned} \tag{3.20}$$

Therefore, it follows from (3.19) and (3.20) that

$$\rho\delta_l^2\widetilde{\alpha}_l^{(k)}\|d_l^{(k)}\|^2 \ge \tau\rho\delta_l^2\widetilde{\alpha}_l^{(k)}\|d_l^{(k)}\|^2 \ge -\widetilde{\mu}_1\nabla\varphi_l^{(k)T}(0)d_l^{(k)}.$$

This implies that

$$\rho\delta_l^2\alpha_l^{(k)} \ge -\widetilde{\mu}_1\|d_l^{(k)}\|^{-2}\nabla\varphi_l^{(k)T}(0)d_l^{(k)}\,,$$

or

$$\alpha_l^{(k)} \ge -\rho^{-1}\widetilde{\mu}_1\delta_l^{-2}\|d_l^{(k)}\|^{-2}\nabla\varphi_l^{(k)T}(0)d_l^{(k)}.$$

We have from (3.15) that

$$\alpha_l^{(k)} \nabla \varphi_l^{(k)T}(0) d_l^{(k)} \le -\rho^{-1} \widetilde{\mu}_1 \delta_l^{-2} \|d_l^{(k)}\|^{-2} (\nabla \varphi_l^{(k)T}(0) d_l^{(k)})^2 \tag{3.21}$$

and

$$(\nabla \varphi_l^{(k)T}(0) d_l^{(k)})^2 \ge \mu_0^2 \|\nabla \varphi_l^{(k)}(0)\|^2 \cdot \|d_l^{(k)}\|^2. \tag{3.22}$$

Letting

$$\omega(\rho,\, \mu_0,\, \widetilde{\mu}_1) = \rho^{-1} \mu_0^2 \widetilde{\mu}_1,$$

$$\sigma_l = \omega(\rho,\, \mu_0,\, \widetilde{\mu}_1) (\max_{1 \le l \le p} \delta_2)^{-2} > 0.$$

Combining (3.16), (3.21) and (3.22), we obtain

$$\begin{aligned} \varphi_l^{(k)}(y_l^{(k)}) - \varphi_l^{(k)}(0) &\le -\omega(\rho,\, \mu_0,\, \widetilde{\mu}_1) \delta_l^{-2} \|\nabla \varphi_l^{(k)}(0)\|^2 + \theta^{(k)} \\ &\le -\omega(\rho,\, \mu_0,\, \widetilde{\mu}_1) (\max_{1 \le l \le p} \delta_l)^{-2} \|\nabla \varphi_l^{(k)}(0)\|^2 + \theta^{(k)} \\ &= -\sigma_l \|\nabla \varphi_l^{(k)}(0)\|^2 + \theta^{(k)}. \quad \square \end{aligned}$$

3.3. A Global Convergence of Algorithm III

To establish the global convergence of Algorithm III, the following assumption is required

(A4)
$$\|A_l^{(k)} y\| \ge \sqrt{\gamma} \|y\|, \quad \forall y \in R^{m_l}. \tag{3.23}$$

Remark. The assumption (A4) is equivalent to saying that the sequence $\{A_l^{(k)T} A_l^{(k)}\}$ of $m_l \times m_l$ matrices is uniformly positive definite, i. e., there exists a constant $\gamma > 0$ independent of k such that for all $y \in R^{m_l}$, $y^T A_l^{(k)T} A_l^{(k)} y \ge \gamma \|y\|^2$.

Lemma 3.2. *If $A_l^{(k)} \in R^{n \times m_l}$ satisfies (A4), then for all $x \in \operatorname{span} A_l^{(k)}$, one has $\|A_l^{(k)T} x\| \ge \sqrt{\gamma} \|x\|$.*

Proof. For all $x \in \operatorname{span} A_l^{(k)}$ we have $x = A_l^{(k)} u$, $\forall u \in R^{m_l}$. Suppose (A4) holds, i. e.,

$$\|A_l^{(k)} y\| \ge \sqrt{\gamma} \|y\|, \quad \forall y \in R^{m_l}. \tag{3.24}$$

Since $A_l^{(k)T} A_l^{(k)} \in R^{m_l \times m_l}$ is nonsingular, (3.24) implies that

$$y^T A_l^{(k)T} A_l^{(k)} y \ge \gamma \|y\|^2 = \beta y^T y,$$

i. e. $y^T[A_l^{(k)T}A_l^{(k)} - \gamma I]y \geq 0$. In consequence, $A_l^{(k)T}A_l^{(k)} - \gamma I \in R^{m_l \times m_l}$ is symmetric positive definite and $A_l^{(k)T}A_l^{(k)}$ is symmetric positive definite as well. It follows that $A_l^{(k)T}A_l^{(k)}[A_l^{(k)T}A_l^{(k)} - \gamma I]$ is symmetric positive definite, i. e., for any $u \in R^{m_l}$, one has $u^T[A_l^{(k)T}A_l^{(k)}]^2 u \geq \gamma u^T A_l^{(k)T}A_l^{(k)} u$, and hence for $x \in \text{span} A_l^{(k)}$ one has $x^T A_l^{(k)} A_l^{(k)T} x \geq \gamma x^T x$, and $\|A_l^{(k)T} x\| \geq \sqrt{\gamma}\|x\|$. □

We turn to establish the convergence of Algorithm III.

Theorem 3.1. *Suppose the following conditions are satisfied:*

(a) $f \in C^1(R^n)$;

(b) $\inf_{x \in R^n} f(x) = \overline{f} > -\infty$;

(c) (A2) and (A4) hold;

(d) $\gamma\sigma_l \geq \lambda$.

Then Algorithm III either terminates at a stationary point of (P) in finite steps or generates an infinite sequence $\{x^{(k)}\}$ *such that*

$$\lim_{k \to \infty} \nabla f(x^{(k)}) = 0,$$

and every accumulation of $\{x^{(k)}\}$ *is a stationary point of (P).*

Proof. First, we assume that Algorithm III terminates, say, at iteration k. Then there exists l_0 such that $\nabla\varphi_{l_0}^{(k)}(0) = 0$, i.e.,

$$\|A_{l_0}^{(k)T} \nabla f(x^{(k)})\| = 0.$$

Since $\nabla f(x^{(k)}) \in \text{span} A_{l_0}^{(k)}$, it follows from (A4) and Lemma 3.2 that $0 = \|A_{l_0}^{(k)T} \nabla f(x^{(k)})\| \geq \sqrt{\gamma}\|\nabla f(x^{(k)})\|$. Thus, one has $\nabla f(x^{(k)}) = 0$. Suppose Algorithm III generates an infinite sequence $\{x^{(k)}\}$, in other words, the algorithm does not terminate in finite steps. We have

$$f(x^{(k)}) - \max_{1 \leq l \leq p} \varphi_l^{(k)}(y_l^{(k)}) \geq \sigma_l \|\nabla\varphi_l^{(k)}(0)\|^2 - \theta_l^{(k)} \quad \text{from (3.12)},$$

$$f(x^{(k)}) - \max_{1 \leq l \leq p} \varphi_l^{(k)}(y_l^{(k)}) \geq \gamma\sigma_l \|\nabla f(x^{(k)})\|^2 - \theta_l^{(k)} \quad \text{from (A4)}$$

and combining (3.13), we have

$$f(x^{(k+1)}) - f(x^{(k)}) \leq -(\gamma\sigma_l - \lambda)\|\nabla f(x^{(k)})\|^2 + \theta_l^{(k)}, \tag{3.25}$$

in terms of (3.13). Without loss of generality, let $\inf_{x\in R^n} f(x) = \overline{f} > 0$, since f is bounded below. Applying (3.25) to the summation of the equality below, we obtain

$$\overline{f} - f(x^{(0)}) \le f(x^{(k)}) - f(x^{(0)}) = \sum_{i=0}^{k-1}(f(x^{(i+1)}) - f(x^{(i)}))$$
$$\le -\sum_{i=0}^{k-1}(\gamma\sigma_l - \lambda)\|\nabla f(x^{(i)})\|^2 + \sum_{i=0}^{k-1}\theta_l^{(i)}.$$

Taking $k \to \infty$, we obtain

$$\overline{f} - f(x^{(0)}) \le -\sum_{i=0}^{\infty}(\gamma\sigma_l - \lambda)\|\nabla f(x^{(i)})\|^2 + \sum_{i=0}^{\infty}\theta_l^{(i)}.$$

According to (A2), we have $\sum_{i=0}^{\infty}(\gamma\sigma_l - \lambda)\|\nabla f(x^{(k)})\|^2 < +\infty$, and hence $\lim_{k\to\infty}(\gamma\sigma_l - \lambda)\|\nabla f(x^{(k)})\|^2 = 0$, i. e., $\lim_{k\to\infty}\nabla f(x^{(k)}) = 0$. Since

$$f(x^{(k+1)}) \le f(x^{(k)}) - (\gamma\sigma_l - \lambda)\|\nabla f(x^{(k)})\|^2 + \theta_l^{(k)} \le f(x^{(k)}) + \theta_l^{(k)},$$

one has that the sequence $\{f(x^{(k)})\}_{k=1}^{\infty}$ converges, according to Lemma 2.1 and (A2). Therefore, for each accumulation $\overline{x}$ of the sequence $\{x^{(k)}\}_{k=1}^{\infty}$, we have $\nabla f(\overline{x}) = 0$. The proof is completed. □

Acknowledgements

The paper is supported by the Young Funds of the State Natural Science No. 10001007; the Scientific Research Foundation of DUT No. 3004888No1.

References

[1] D.P. Bertsekas, J.N. Tsitsiklis, *Parallel, Distributed Computation: Numerical Methods*, Prentice-Hall: Englewood Cliffs, New Jersey (1989).

[2] Y.C. Cheng, On the gradient projection method for solving the nonsymmetric linear complementarity problem, *Journal of Optimization Theory, Applications*, **43** (1984), 527-541.

[3] M.C. Ferris, O.L. Mangasarian, Parallel variable distribution, *SIAM J. on Optimization*, **4** (1994), 102-126.

[4] M. Fukushima, Parallel variable transformstion in unconstrained optimization, *SIAM J. on Optimization*, **8** (1998), 658-672.

[5] D. Goldfarb, Curvilinear path steplength algorithms for minimization which use directions of negative curvature, *Mathematical Programming*, **18** (1980), 31-40.

[6] O.L. Mangasarian, Parallel gradient distribution in unconstrained optimization, *SIAM J. Control Optim.*, **33** (1995), 1916-1925.

[7] O.L. Mangasarian, M.V. Solodov, Serial, parallel backpropagation convergence via nonmonotone perturbed minimization, *Optimization Methods in Software*, **4** (1994), 103-116.

[8] J.M. Ortega, W.C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York (1970).

[9] J.S. Pang, S.P. Han, N. Pangaraj, Mininization of locally Lipschitzian functions, *SIAM J. Optimization*, **1** (1991), 57-82.

[10] M.V. Solodov, New inexact parallel variable distribution algorithms, *Computational Optimization, Applications*, **7** (1997), 165-182.

[11] M.V. Solodov, S.K. Zavriev, Error-stabilty properties of generalized gradient-type algorithm, *Mathematical Programming Technica Report*, 94-05, Computer Science Department, University of Wisconsin, 1210 West Dayton Street, Madison, Wisconsin 53706, U.S.A. (June 1994), Revised: July 1995.