Scientific Research

A New Inexact PVD Algorithm for General Programming Problem

XU Fangfang, LIU Weihui, Zhu Gengfeng

College of Information science and Technology, Shandong University of Science and Technology, Qingdao, China, 266510

e-mail: happyxufangfang@126.com

Abstract: This paper consider the parallel variable distribution (PVD) approach proposed by Ferris and Mangasarian for solving optimization problem. This paper propose to apply the PVD approach to problems with general convex constraints and show that the algorithm converges, provided certain conditions are imposed on the change of secondary variables. In this paper, we choose to use the projected gradient direction for secondary variables and replace the minimization problem with a sufficient descent condition in the parallelization stage.

Keywords: parallel variable distribution; projected gradient residual function; sufficient descent condition; general convex constraints

1 Introduction

We first describe the PVD algorithm proposed in [1].We consider the nonlinear program: $\min_{x\in X} f(x)$ where X is a nonempty closed convex set in R^n and $f: R^n \to R$ has continuous first partial derivatives in R^n. Suppose there are p processors which divide the variable $x\in R^n$ into blocks $x_1, x_2, \ldots, x_p$, where $x^l \in n^l$, $\sum_{l=1}^{p} n^l = l$, among p processors. Given iteration $x^i \in R^n$, processor l has primary responsibility for updating block $x_l^i \in R^{n_l}$ of the iterate x^i. The general framework of PVD algorithm is as follows:

Algorithm 1. Start with any $x^0 \in R^n$. Having x^i, stop if $\nabla f\left(x^i\right) = 0$. Otherwise, compute x^{i+1} as follows:

- Parallelization: For each processor $l \in \{1,\ldots,p\}$

(1.1) compute $\left(y_l^i, \mu_{\bar{l}}^i\right)$

$$\in \arg\min_{x_l,\mu_{\bar{l}}} \psi_l^i\left(x_l, \mu_{\bar{l}}\right) := f\left(x_l, x_{\bar{l}}^i + D_{\bar{l}}^i \mu_{\bar{l}}\right)$$

- Synchronization: compute x^{i+1} such that

$$f\left(x^{i+1}\right) \le \min_{l=\{1,\ldots,p\}} \psi_l^i\left(y_l^i, \mu_{\bar{l}}^i\right)$$

We will sometimes refer to x^i as the base point at the $(i+1)$-st iteration. In the above algorithm $\bar{l}$ denotes the complement of l in the set $\{1,\ldots,p\}$ and $\mu_{\bar{l}} \in R^{p-1}$. The matrix $D_{\bar{l}}^i$ is an $n_{\bar{l}} \times (p-1)$ block diagonal matrix formed by placing the blocks $d_1^i,\ldots,d_{p-1}^i$ ($d_t^i \in R^{n_t}$, $t = 1,\ldots,p-1$) of an arbitrary direction $d^i \in R^n$ along its block diagonal.

Foundation Item: The Science and Technology Plan of Shandong Province (Y2008A01)

We consider in this paper the following problem[2]:

(1.2) $\min f(x)$ such that $g(x) \le 0$

where $f: R^n \to R$ and $g: R^n \to R^m$ are $C^1\left(R^n\right)$. If the constraints are not separable, having a stationary point that results from minimizing the objective function with respect to individual blocks of variables $x_1, x_2, \ldots, x_p$ and subject to the problem constraints $g(x) \le 0$ does not result in a useful point. This is easily illustrated by the following simple example in R^2:

(1.3) $\min x_1^2 + x_2^2$ such that $x_1 + x_2 \ge 2$

This strongly convex problem has a unique global solution at $x_1 = x_2 = 1$. However, the point $x_1 = 0.5, x_2 = 1.5$ is a global minimum with respect to each of x_1 and x_2 separately, subject to the problem constraint. That is

$$0.5 = \arg\min_{x_1} \left\{x_1^2 + 2.25 \middle| x_1 \ge 0.5\right\}$$

$$1.5 = \arg\min_{x_2} \left\{0.25 + x_2^2 \middle| x_2 \ge 1.5\right\}$$

This property is possessed by all points in the first quadrant lying on the constraint $x_1 + x_2 = 2$. None of these points are of any use except the solution $x_1 = x_2 = 1$.

2 PVD with inexact subproblem solution

In this paper, we choose projected gradient residual function[2]:

(2.1) $r(x) := x - P_c\left[x - \nabla f(x)\right]$

as PVD direction.

(2.2) $d_l^i = r_l\left(x^i\right) := \left[r\left(x^i\right)\right]_l \; l = 1,\cdots p$

Furthermore, we replace the minimization problem in (1.1) with the following sufficient descent condition:

(2.3) Computer $\left(y_l^i, \mu_{\bar{l}}^i\right)$ such that

$$\psi_l^i\left(y_l^i, \mu_{\bar{l}}^i\right) \le \psi_l^i\left(x_l - \eta_i r_l\left(x^i\right), -\eta_i e_{\bar{l}}\right)$$

where $e_{\bar{l}}$ is a vector of ones of appropriate dimension, and (2.4):

$$\eta_i := \min\left\{1, \beta\left(\frac{1+\alpha}{L}\right)^{1/\alpha}\left\|r_l\left(x^i\right)\right\|^{(1-\alpha)/\alpha}\right\} \quad \beta\in(0,1)$$

So we can get a new inexact PVD algorithm for problem (1.2).

Algorithm2.5. Start with any $x^0 \in R^n$. Having x^i, stop if $\nabla f\left(x^i\right) = 0$. Otherwise, compute x^{i+1} as follows:

- Parallelization: For each Processor $l \in \{1,\dots,p\}$

$$\min_{x_l,\mu_{\bar{l}}} \psi_l^i\left(x_l, \mu_{\bar{l}}\right) := f\left(x_l, x_{\bar{l}}^i + D_{\bar{l}}^i \mu_{\bar{l}}\right)$$

Computer $\left(y_l^i, \mu_{\bar{l}}^i\right)$ such that

$$\psi_l^i\left(y_l^i, \mu_{\bar{l}}^i\right) \le \psi_l^i\left(x_l - \eta_i r_l\left(x^i\right), -\eta_i e_{\bar{l}}\right)$$

- Synchronization: compute x^{i+1} such that

$$f\left(x^{i+1}\right) \le \min_{l=\{1,\dots,p\}} \psi_l^i\left(y_l^i, \mu_{\bar{l}}^i\right)$$

Theorem 2.6. Let $f(\cdot) \in C_L^{1,\alpha}\left(R^n\right)$. Suppose $\left\{x^i\right\}$ is any sequence generated by PVD algorithm 2.2. Then either $f(\cdot)$ is unbounded from below on C or the sequence $\left\{f\left(x^i\right)\right\}$ converges, the sequence $\left\{r\left(x^i\right)\right\}$ converges to zero and every accumulation point of the sequence $\left\{x^i\right\}$ satisfies the minimum principle necessary optimality condition.

Proof:

$$\begin{aligned}
&\left(x_l - \eta_i r_l\left(x^i\right), x_{\bar{l}} - \eta_i D_{\bar{l}}^i e_{\bar{l}}\right)\\
&= \left(x_l - \eta_i r_l\left(x^i\right), x_{\bar{l}} - \eta_i r_{\bar{l}}\left(x^i\right)\right)\\
&= x^i - \eta_i r\left(x^i\right)\\
&= \left(1-\eta_i\right)x^i + \eta_i P_C\left[x^i - \nabla f\left(x^i\right)\right] \in C
\end{aligned}$$

where the first equality follows from the block diagonal structure of $D_{\bar{l}}^i$ and (2.2), the last equality follows from (2.1), and the inclusion is by convexity of the set C. We further obtain

$$\begin{aligned}
&f\left(x^i\right) - f\left(y_l^i, x_{\bar{l}} + D_{\bar{l}}^i \mu_{\bar{l}}^i\right)\\
&= \psi_l^i\left(x_l^i, 0\right) - \psi_l^i\left(y_l^i, \mu_{\bar{l}}^i\right)\\
&\ge \psi_l^i\left(x_l^i, 0\right) - \psi_l^i\left(x_l - \eta_i r_l\left(x^i\right), -\eta_i e_{\bar{l}}\right)\\
&= f\left(x^i\right) - f\left(x_l - \eta_i r_l\left(x^i\right), x_{\bar{l}} - \eta_i D_{\bar{l}}^i e_{\bar{l}}\right)\\
&= f\left(x^i\right) - f\left(x^i - \eta_i r\left(x^i\right)\right)
\end{aligned}$$

where the inequality follows from (2.3). Furthermore,

$$\begin{aligned}
&f\left(x^i - \eta_i r\left(x^i\right)\right) - f\left(x^i\right)\\
&= -\int_0^{\eta_i}\left\langle \nabla f\left(x^i - t r\left(x^i\right)\right), r\left(x^i\right)\right\rangle dt\\
&= -\eta_i\left\langle \nabla f\left(x^i\right), r\left(x^i\right)\right\rangle\\
&\quad -\int_0^{\eta_i}\left\langle \nabla f\left(x^i - t r\left(x^i\right)\right) - \nabla f\left(x^i\right), r\left(x^i\right)\right\rangle dt\\
&\le -\eta_i\left\langle \nabla f\left(x^i\right), r\left(x^i\right)\right\rangle + L\int_0^{\eta_i} t^\alpha \left\|r\left(x^i\right)\right\|^{1+\alpha} dt\\
&= -\eta_i\left\langle \nabla f\left(x^i\right), r\left(x^i\right)\right\rangle + \frac{L\eta_i^{\alpha+1}}{1+\alpha}\left\|r\left(x^i\right)\right\|^{1+\alpha}
\end{aligned}$$

where the equality follows from the Holder continuity of $\nabla f(\cdot)$ and the Cauchy-Schwarz inequality. Hence

(2.7) $f\left(x^i\right) - f\left(y_l^i, x_{\bar{l}} + D_{\bar{l}}^i \mu_{\bar{l}}^i\right)$

$$\ge \eta_i\left\langle \nabla f\left(x^i\right), r\left(x^i\right)\right\rangle - \frac{L\eta_i^{\alpha+1}}{1+\alpha}\left\|r\left(x^i\right)\right\|^{1+\alpha}$$

By properties of the projection operator[5], for any $x \in R^n$ and any $y \in C$, it holds that

$$\left\langle x - P_C(x), y - P_C(x)\right\rangle \le 0$$

Taking $x = x^i - \nabla f\left(x^i\right)$ and $y = x^i \in C$, we have:

$$\begin{aligned}
0 &\ge \left\langle x^i - \nabla f\left(x^i\right) - P_C\left(x^i - \nabla f\left(x^i\right)\right), x^i - P_C\left(x^i - \nabla f\left(x^i\right)\right)\right\rangle\\
&= \left\langle r\left(x^i\right) - \nabla f\left(x^i\right), r\left(x^i\right)\right\rangle
\end{aligned}$$

Hence $\left\langle r\left(x^i\right), r\left(x^i\right)\right\rangle \le \left\langle \nabla f\left(x^i\right), r\left(x^i\right)\right\rangle$

The latter relation combined with (2.7) yields

$$\begin{aligned}
&f\left(x^i\right) - f\left(y_l^i, x_{\bar{l}} + D_{\bar{l}}^i \mu_{\bar{l}}^i\right)\\
&\ge \eta_i\left\langle r\left(x^i\right), r\left(x^i\right)\right\rangle - \frac{L\eta_i^{\alpha+1}}{1+\alpha}\left\|r\left(x^i\right)\right\|^{1+\alpha}\\
&= \eta_i\left\|r\left(x^i\right)\right\|^2\left(1 - \frac{L\eta_i^\alpha}{1+\alpha}\left\|r\left(x^i\right)\right\|^{\alpha-1}\right) \ge \left(1-\beta^\alpha\right)\eta_i\left\|r\left(x^i\right)\right\|^2\\
&\ge \left(1-\beta^\alpha\right)\min\\
&\left\{\left\|r\left(x^i\right)\right\|^2, \beta\left(\frac{1+\alpha}{L}\right)^{1/\alpha}\left\|r_l\left(x^i\right)\right\|^{(1+\alpha)/\alpha}\right\}
\end{aligned}$$

where the last two inequalities follow from the choice of η_i. By the synchronization step , we have

(2.8) $f\left(x^i\right) - f\left(x^{i+1}\right) \ge$

$$\left(1-\beta^{\alpha}\right)\min\left\{\left\|r\left(x^{i}\right)\right\|^{2},\beta\left(\frac{1+\alpha}{L}\right)^{1/\alpha}\left\|r_{l}\left(x^{i}\right)\right\|^{(1+\alpha)/\alpha}\right\}$$

Thus the sequence $\left\{f\left(x^{i}\right)\right\}$ is nonincreasing. If $f(\cdot)$ is bounded below on C, then $\left\{f\left(x^{i}\right)\right\}$ is bounded and hence it converges. In the latter case, $\left\{f\left(x^{i}\right)-f\left(x^{i+1}\right)\right\}\to 0$ and therefore $\left\{r\left(x^{i}\right)\right\}\to 0$ by (2.8). By continuity of $r\left(x^{i}\right)$, it follows that for every accumulation point $\bar{x}$ of the sequence $\left\{x^{i}\right\}$, $r\left(\bar{x}\right)=0$. Thus all accumulation points of $\left\{x^{i}\right\}$ satisfy the minimum principle necessary optimality condition.

Consider the problem (1.3) above, start with the point $x_1^0=0.5$, $x_2^0=1.5$. It is easy to check that $r_1\left(x^0\right)=-1$, $r_2\left(x^0\right)=-1$. Thus the two parallel subproblems in the PVD algorithm are

$$\min_{x_1,\mu_2} f\left(x_1,x_2^0+\mu_2 r_2\left(x^0\right)\right)=\left(x_1\right)^2+\left(\mu_2\right)^2+2x_2^0\mu_2+\left(x_2^0\right)^2$$

$$s.t.\qquad 0\le g\left(x_1,x_2^0+\mu_2 r_2\left(x^0\right)\right)=x_1+\mu_2-0.5$$

and

$$\min_{x_2,\mu_1} f\left(x_1^0+\mu_1 r_1\left(x^0\right),x_2\right)=-2x_1^0\mu_1+\left(\mu_1\right)^2+\left(x_2\right)^2+\left(x_1^0\right)^2$$

$$s.t.\qquad 0\le g\left(x_1^0+\mu_1 r_1\left(x^0\right),x_2\right)=x_2-\mu_1-1.5$$

The solution of the first subproblem is $y_1^0=1$, $\mu_2^0=-0.5$, and the solution of the second is $y_2^0=1$, $\mu_1^0=-0.5$. It is easy to see that $x_1^0+\mu_1^0 r_1$, $\left(x^0\right)=1$ $x_2^0+\mu_2^0 r_2\left(x^0\right)=1$. So both subproblems produce the solution of the original problem.

3 Conclusions

It was established that the parallel variable distribution approach can be successfully applied to solve optimization problems with general convex constraints. Furthermore , conditions imposed on solving parallel subproblems were considerably relaxed, thus yielding a more practical framework.

4 Acknowledgment

First and foremost, I am in considerable debt to my supervisor, Prof. HE Gou-ping, who gave me inspiring encouragement, invaluable advice, and strict instruction. I am also grateful to the sponsorship from The Science and Technology Plan of Shandong Province (Y2008A01).

Referemces

[1] M. C. Ferris and O.L. Mangasarian. Parallel variable distribution. SIAM J. OPTIM. 1994. 4(4): 815-832.

[2] Michael. V. Solodov. On The Convergence of Constrained Parallel Variable distribution Algorithms. SIAM J. OPTIM. 1998.8(1):187-196;

[3] Lai Yanlian and He Guoping. Optimization methods. Tsing Hua University Press. Beijing, 2008. 313-359.

[4] Michael. V. Solodov. New inexact parallel variable distribution algorithms. Computational Optimization and Applications. 1997. 7(2):165-182.

[5] C.A.Sagastizabal, M.V.Solodov. Parallel variable distribution for constrained optimization. Computational optimization and applications. 2002(22): 111~13.

978-1-935068-06-8 © 2009 SciRes.