大規模非線形計画問題に対する逐次二次計画分解法*

若原達朗**・福島雅夫***・茨木俊秀***

A Successive Quadratic Programming Decomposition Method for Large-Scale Nonlinear Programming Problems*

Tatsuro WAKAHARA**, Masao FUKUSHIMA*** and Toshihide IBARAKI***

Large-scale nonlinear programming problems usually contain a relatively small number of nonlinear variables. For such a problem, it is often effective to treat the linear part as a linear programming problem by temporarily fixing the nonlinear variables. As a result, we obtain a nonsmooth optimization problem containing the nonlinear variables only. Based on this idea, we propose a successive quadratic programming algorithm for large-scale nonlinear programming problems. We show convergence of the algorithm and report some computational experience.

1. はじめに

実際に現れる大規模な非線形計画問題では、目的関数、制約条件の双方とも、大部分の変数に対して線形であり、非線形変数はごくわずかであることが少なくない。またこのとき、線形変数はいくつかのブロックに分かれて現れることが多い。非線形変数を $x \in R^n$, 第 i ブロックの線形変数を $y_i \in R^{n_i}$, $i = 1, \dots, l$, とすると、このような問題は次のように表わされる。

$$\begin{aligned} P: \text{minimize } & F(x) + \sum_{i=1}^l c_i^T y_i \\ \text{subject to } & A_i^T x + B_i^T y_i \leq b_i, \quad i = 1, \dots, l \\ & g_j(x) \leq 0, \quad j = 1, \dots, m. \end{aligned}$$

ただし、 $F: R^n \rightarrow R$ と $g_j: R^n \rightarrow R$ は 2 回連続微分可能な関数、 A_i, B_i, b_i, c_i はそれぞれ $n \times m_i, n_i \times m_i, m_i, n_i$ 次元の行列やベクトルとし、 T は転置を表わすものとする。関数 F と g_j は一般の非線形関数であり、凸であるとは限らない。このような構造をもつ問題として定式化される典型的な問題としては、たとえば 2 段階確率計画問題がある。(参考文献 1) の 2.5 節や参考文献 2) の 6.6 節に、線形確率計画問題を問題 P の形 (ただし、

F と g_j は線形) に定式化する方法が示されている。元の確率計画問題が非線形ならば、同様の定式化によって得られる問題は上の問題 P の形になる。)

本論文の目的は、問題 P に対してその構造を利用した新しい分解アルゴリズムを提案することである。問題 P は、非線形変数 x を一時的に固定すると、変数 y_i のみを含む l 個の独立な線形計画問題に分解できるという構造を持っており、よく知られた Benders の分解法^{1)~3)} はそのような性質を利用している。本論文ではまず、正確なペナルティ関数 (exact penalty function)^{4), 5)} を用いて問題 P を再定式化した後、それを線形変数 y_i に関して最適化した結果得られる非線形計画問題に対して逐次二次計画法^{6), 7)} を適用する。とくに、この方法では各反復において二次計画部分問題を解くが、それらの部分問題の制約条件の数は $m+2$ を越えることはない。このことは、部分問題の制約条件の数が単調に増加していく Benders の方法と比べて、記憶量の面で有利であることを示している。また、通常の Benders の方法^{1)~3)} では、各反復で部分問題の厳密解を求める必要があるが、これは例えば関数 F や g_j がすべて線形であるような特別な場合を除いて困難である。しかし本方法においては、部分問題は凸二次計画問題であるから、その厳密解を常に有限ステップで計算できるという利点がある。

以下では $g(x) = (g_1(x), \dots, g_m(x))^T$, $G(x) = [\nabla g_1(x), \dots, \nabla g_m(x)]^T$ と書くこととし、 $\|\cdot\|$ と $\|\cdot\|_\infty$ はそれぞれ l_2 ノルムと l_∞ ノルムを、また凸関数 $f(x)$ の劣微分⁸⁾ を $\partial f(x)$ で表わすものとする。

* 原稿受付 1990. 12. 10

** 東邦ガス (株) Toho Gas Co., Ltd.; Sakurada-cho 19-18, Atsuta-ku, Nagoya-city, Nagoya 456, JAPAN.

*** 京都大学 工学部 Faculty of Engineering, Kyoto University; Yoshida-Honmachi, Sakyo-ku, Kyoto 606, JAPAN

Key Words: decomposition, successive quadratic programming, large-scale problem, nonsmooth optimization, exact penalty function.

2. 問題の分解と定式化

問題 P において変数 x を固定すると、次の各 y_i に関する l 個の独立な線形計画問題 $P_i(x)$ が得られる。

$$P_i(x): \text{minimize } c_i^T y_i \\ \text{subject to } B_i^T y_i \leq b_i - A_i^T x.$$

この問題はすべての $x \in R^n$ に対して実行可能というわけではない。そこで問題 $P_i(x)$ が実行可能であるような x の集合を

$$\Omega_i = \{x \mid \exists y_i \in R^{m_i} \text{ such that } A_i^T x + B_i^T y_i \leq b_i\}$$

で定義し、さらに $M_i > 0$ を定数、 $w_i \in R^{m_i}$ 、 $u_i \in R^{m_i}$ を変数ベクトル、 $e_i = (1, 1, \dots, 1)^T \in R^{m_i}$ として、次の二つの双対な線形計画問題の組

$$\begin{aligned} \hat{P}_i(x): \text{minimize } c_i^T y_i + M_i e_i^T w_i \\ \text{subject to } B_i^T y_i - w_i \leq b_i - A_i^T x \\ w_i \geq 0 \\ \hat{D}_i(x): \text{maximize } (A_i^T x - b_i)^T u_i \\ \text{subject to } B_i u_i = -c_i \\ 0 \leq u_i \leq M_i e_i \end{aligned}$$

を考える。 $\hat{P}_i(x)$ はすべての $x \in R^n$ に対して実行可能であり、また、 $\hat{D}_i(x)$ の制約集合は有界かつ x とは独立であるから、この集合が空でないときは、 $\hat{D}_i(x)$ はすべての $x \in R^n$ に対して有限な解を持つ。さらに $\hat{P}_i(x)$ の最適解を $(\bar{y}_i, \bar{w}_i)$ 、 $\hat{D}_i(x)$ の最適解を $\bar{u}_i$ とすると、LP の双対定理により、 $c_i^T \bar{y}_i + M_i e_i^T \bar{w}_i = (A_i^T x - b_i)^T \bar{u}_i$ であるので、これと $B_i \bar{u}_i = -c_i$ から、

$$(A_i^T x + B_i^T \bar{y}_i - b_i)^T \bar{u}_i = M_i e_i^T \bar{w}_i$$

が成り立つ。

問題 $P_i(x)$ と問題 $\hat{P}_i(x)$ 、 $\hat{D}_i(x)$ のあいだには次の関係が成り立つ⁵⁾。

- ① $x \notin \Omega_i$ のとき、
  - (a) $P_i(x)$ は実行可能でない。
  - (b) $\hat{P}_i(x)$ は最適解 $(\bar{y}_i, \bar{w}_i)$ をもち、 $\hat{D}_i(x)$ も最適解 $\bar{u}_i$ を持つ。ここで、 $\bar{w}_i = \max\{0, A_i^T x + B_i^T \bar{y}_i - b_i\}$ (ただし $\max$ は要素ごとにとるものとする) であり、 $e_i^T \bar{w}_i > 0$ 、すなわち $(A_i^T x + B_i^T \bar{y}_i - b_i)^T \bar{u}_i > 0$ である。
- ② $x \in \Omega_i$ であるが、 $P_i(x)$ が有限な解を持たないとき、 $\hat{P}_i(x)$ もまた有限な解を持たない。さらに $\hat{D}_i(x)$ は実行可能でない。この場合はもとの問題 P も有限な解を持たない。
- ③ $x \in \Omega_i$ であり、かつ $P_i(x)$ が最適解 $\bar{y}_i$ を持つとき、 $\hat{P}_i(x)$ 十分大きい M_i に対して最適解 $(\bar{y}_i, 0)$ を持つ。このとき、 $\hat{D}_i(x)$ も最適解 $\bar{u}_i$ を持ち、

$(A_i^T x + B_i^T \bar{y}_i - b_i)^T \bar{u}_i = 0$ が成り立つ。

② の場合を除いて $\hat{P}_i(x)$ と $\hat{D}_i(x)$ はともに最適解を持つ。このとき LP の双対定理により二つの問題の最適値は一致するが、この最適値は x に依存している。そこで x の関数 $f_i: R^n \rightarrow R$ を

$$\begin{aligned} f_i(x) &= \min \{c_i^T y_i + M_i e_i^T w_i \mid B_i^T y_i - w_i \\ &\leq b_i - A_i^T x, w_i \geq 0\} \\ &= \max \{(A_i^T x - b_i)^T u_i \mid B_i u_i = -c_i, \\ &0 \leq u_i \leq M_i e_i\} \end{aligned}$$

で定義し、さらに

$$f(x) = \sum_{i=1}^l f_i(x)$$

によって関数 $f: R^n \rightarrow R$ を定義する。関数 f は区分的に線形な凸関数である。

以上の考察により、問題 P が最適解を持つならば、定数 $M_i > 0$ 、 $i = 1, \dots, l$ が十分大きいとき、問題 P と次の問題 $\hat{P}$ の最適解は一致する。

$$\begin{aligned} \hat{P}: \text{minimize } F(x) + f(x) \\ \text{subject to } g(x) \leq 0 \end{aligned}$$

とくに、この問題は次の問題と等価であることに注意しておこう。

$$\begin{aligned} \text{minimize } F(x) + \sum_{i=1}^l (c_i^T y_i + M_i e_i^T w_i) \\ \text{subject to } A_i^T x + B_i^T y_i - w_i \leq b_i, i = 1, \dots, l \\ g(x) \leq 0, w \geq 0 \end{aligned}$$

問題 P の変数が $x, y_1, \dots, y_l$ であるのに対し、問題 $\hat{P}$ の変数は x のみであり、 x の次元にくらべて各 y_i の次元が相当大きいときには、 P の代わりに $\hat{P}$ を扱うことで問題の規模を格段に小さくすることができる。さらに、各 $f_i(x)$ を定義する線形計画問題 $\hat{P}_i(x)$ は互いに独立であり、並列処理も可能である。ただし、 $f(x)$ は区分的に線形な凸関数であり、必ずしも微分可能とは限らないので、問題 $\hat{P}$ は微分不可能な最適化問題となっている。

3. アルゴリズム

問題 $\hat{P}$ は、最近参考文献6)で提案されたアルゴリズムの考え方をを用いて解くことができる。このアルゴリズムでは、パラメータ $r > 0$ を含む正確なペナルティ関数

$$W(x, r) = f(x) + F(x) + r \sum_{i=1}^m \max\{0, g_i(x)\}$$

を定義し、さらに固定された x に対する部分問題として、 $p \in R^n$ を変数とする次の問題 $SP(x)$ を考える。

$$\begin{aligned} SP(x): \text{minimize } \Phi(p) \\ \text{subject to } g(x) + G(x)p \leq 0 \end{aligned}$$

ただし、関数 Φ は適当な正定値対称行列 H を用いて次式で定義する。

$$\Phi(p) = f(x+p) + F(x) + \nabla F(x)^T p + \frac{1}{2} p^T H p$$

【定理 1】

問題 $SP(x)$ の最適解とそれに対応するラグランジュ乗数の組を $(\bar{p}, \bar{v})$ とする。 $\bar{p} = 0$ ならば $(x, \bar{v})$ は問題 $\hat{P}$ の Kuhn-Tucker 条件を満足する。 $\bar{p} \neq 0$ ならば、 $r \geq \|\bar{v}\|_\infty$ であるような r に対して、 $\bar{p}$ はペナルティ関数 $W(x, r)$ の点 x における降下方向である。

証明 付録 1 参照。

問題 $SP(x)$ は微分不可能な凸関数 f を含んでいるので、その厳密解を計算することは一般に容易ではない。参考文献 6) のアルゴリズムでは、凸関数 f に対して逐次生成した線形近似関数を蓄積していく、いわゆる切除平面法の考え方をを用いて得られた二次計画問題を反復して解くことにより、問題 $SP(x)$ の近似解を求めている。しかしながら、その方法には反復が進むにつれて、解くべき二次計画問題の制約条件の数が単調に増加するという難点がある。そこで、以下では参考文献 7) のアルゴリズムの考え方を取り入れ、過去の反復において生成された線形近似関数、すなわち“切除平面”を常に一つの切除平面に集約することにより、記憶量の大幅な削減を試みる。

まず p^0 を任意に選び、 $j = 1, 2, \dots$ に対して次の問題 $QP_j(x)$ の最適解 p^j を求める。

$$\begin{aligned} QP_j(x): \text{minimize } \Psi_j(p) \\ \text{subject to } g(x) + G(x)p \leq 0 \end{aligned}$$

ただし、関数 $\Psi_j: R^n \rightarrow R$ は

$$\begin{aligned} \Psi_j(p) = \max\{h_{j-1}(p), h_{-(j-1)}(p)\} + F(x) \\ + \nabla F(x)^T p + \frac{1}{2} p^T H p \end{aligned}$$

で定義されるものとし、さらに関数 $h_j, j = 0, 1, \dots$, および $h_{-j}, j = 1, 2, \dots$, は、 $\xi^j \in \partial f(x+p^j)$ を用いて

$$\begin{aligned} h_j(p) &= f(x+p^j) + (\xi^j)^T (p-p^j) \\ h_{-j}(p) &= \lambda_j h_{j-1}(p) + (1-\lambda_j) h_{-(j-1)}(p) \end{aligned}$$

で定義する。ここで、 λ_j は問題 $QP_j(x)$ と等価な問題

$$\begin{aligned} \text{minimize } z + F(x) + \nabla F(x)^T p + \frac{1}{2} p^T H p \\ \text{subject to } h_l(p) - z \leq 0, l = j-1, -(j-1) \\ g(x) + G(x)p \leq 0 \end{aligned}$$

において、制約条件 $h_{j-1}(p) - z \leq 0$ に対応する最適ラグランジュ乗数であり、 $0 \leq \lambda_j \leq 1$ である。上の関数 h_j は、第 j 部分問題 $QP_j(x)$ の解によって生成される切除平面に対応しており、関数 h_{-j} は第 j 部分問題以前

に生成された切除平面を集約したものである。

【定理 2】

$\bar{p}$ と $p^j, j = 1, 2, \dots$, をそれぞれ問題 $SP(x), QP_j(x)$ の最適解、 v^j を $QP_j(x)$ の最適ラグランジュ乗数、 Λ を問題 $SP(x)$ の最適ラグランジュ乗数の集合とする。また問題 $SP(x)$ は一次独立制約想定⁸⁾をみたすものとする。このとき $\{p^j\}$ と $\{v^j\}$ が有界ならば次が成り立つ。

- (1) $\Psi_j(p^j) \leq \Psi_{j+1}(p^{j+1}) \leq \Phi(\bar{p}), j = 1, 2, \dots$
- (2) $\Psi_j(p^j) \rightarrow \Phi(\bar{p})$
- (3) $\Phi(p^j) \rightarrow \Phi(\bar{p})$
- (4) $\|p^j - \bar{p}\| \rightarrow 0$
- (5) $\inf\{\|v^j - v\| | v \in \Lambda\} \rightarrow 0$

証明 付録 2 参照。

以上の考察に基づいて、問題 $\hat{P}$ に対するアルゴリズムを以下のように構成することができる。このアルゴリズムでは、七つのパラメータ $\gamma \in (0, \infty), \delta \in (0, \infty), \rho \in (0, \infty), \xi \in (1, \infty), \beta \in (0, 1), \sigma \in (0, 1), \tau \in (0, 1)$ を使用する。また、任意の点 x において凸関数 $f(x)$ の少なくとも一つの劣勾配 $\xi \in \partial f(x)$ を求める必要があるが、これは問題 $\hat{D}_i(x)$ の最適解 $\bar{u}_i$ を用いて、 $\xi = \sum_{i=1}^l A_i \bar{u}_i$ とおくことにより得られる。

アルゴリズム SQPD

Step 0: 初期点 $x^1 \in R^n$ と $\bar{r}_0 > 0$ を選び、 $k := 1$ とする。

Step 1: 正定値対称行列 H_k を選び、関数 Φ_k を次で定義する。

$$\begin{aligned} \Phi_k(p) &= f(x^k+p) + F(x^k) + \nabla F(x^k)^T p \\ &\quad + \frac{1}{2} p^T H_k p \end{aligned}$$

p^0 を選び、 $f(x^k+p^0)$ と $\xi^0 \in \partial f(x^k+p^0)$ を問題 $\hat{D}_i(x^k+p^0), i = 1, \dots, l$, を解くことにより求める。 $h_0(p) = f(x^k+p^0) + \xi^{0T}(p-p^0)$ とおき、 $r_0 := \bar{r}_{k-1}, j := 1$ とする。

Step 2: 関数 Ψ_{kj} を次で定義する。

$$\begin{aligned} \Psi_{kj}(p) &= \max\{h_{j-1}(p), h_{-(j-1)}(p)\} + F(x^k) \\ &\quad + \nabla F(x^k)^T p + \frac{1}{2} p^T H_k p \end{aligned}$$

問題 $QP_{kj}: \min\{\Psi_{kj}(p) | g(x^k) + G(x^k)p \leq 0\}$ と等価な二次計画問題

$$\begin{aligned} \text{minimize } z + F(x^k) + \nabla F(x^k)^T p \\ + \frac{1}{2} p^T H_k p \end{aligned}$$

subject to $h_l(p) - z \leq 0, l = j-1, -(j-1)$
 $g(x^k) + G(x^k)p \leq 0$

の最適解 p^j と、制約条件 $h_{j-1}(p) - z \leq 0$ に対

応するラグランジュ乗数 λ_j , $g(x^k) + G(x^k)p \leq 0$ に対応するラグランジュ乗数 v^j を求める。さらに $\xi^i \geq \|v^j\|_\infty + \rho$ を満たす最小の非負整数 i を i_j とし, $r_j = \xi^{i_j} r_{j-1}$ とする。

Step 3: 次の条件が成り立つならば Step 4 へ; そうでなければ Step 6 へ。

$$W(x^k, r_j) \geq \Phi_k(p^j) - \gamma$$

Step 4: 次の条件が満たされれば終了; そうでなければ Step 5 へ。

$$W(x^k, r_j) \leq \Psi_{kj}(p^j) + \delta$$

Step 5: 次の条件が成り立つならば, $d^k := p^j$, $\bar{r} := r_j$ として Step 7 へ; そうでなければ Step 6 へ。

$$\Phi_k(p^j) - \Psi_{kj}(p^j) \leq \beta \{W(x^k, r_j) - \Psi_{kj}(p^j)\}$$

Step 6: 問題 $\hat{D}_i(x^k + p^j)$, $i = 1, \dots, l$, を解いて $\xi^j \in \partial f(x^k + p^j)$ を求める。 $h_j(p) = f(x^k + p^j) + (\xi^j)^T(p - p^j)$, $h_{-j}(p) = \lambda_j h_{-(j-1)}(p) + (1 - \lambda_j) h_{-(j-1)}(p)$, $j := j+1$ として Step 2 に戻る。

Step 7: 次の条件を満足する最小の非負の整数 l を l_k とする。

$$W(x^k + \tau^l d^k, \bar{r}_k) \leq W(x^k, \bar{r}_k) - \sigma q_k \tau^l$$

ただし $q_k = (1 - \beta)\delta + \frac{1}{2}(d^k)^T H_k d^k$ である。

$\alpha_k := \tau^{l_k}$, $x^{k+1} := x^k + \alpha_k d^k$, $k := k+1$ として Step 1 へ戻る。

上のアルゴリズムの Step 1 における行列 H_k の選び方についてはさまざまな方法が考えられる。最も簡単な方法は, H_k を常に単位行列と置く方法であり, これは最急降下法の一種の拡張とみなせる。また, 問題 $\hat{P}$ のラグランジュ関数のヘッセ行列を近似するように, 準ニュートン法の公式を用いて逐次更新することもできる。ただし, この問題の場合, ラグランジュ関数は微分可能ではないので, ヘッセ行列や勾配は一般に存在しない。したがって, 勾配のかわりに劣勾配を用いて, 準ニュートン法を形式的に模倣することになる。なお, アルゴリズムの収束性を保証するには, 以下の定理に示すように, 行列 H_k が一様に正定値かつ有界に保たれれば十分である。

次の定理は, 適当な仮定のもとで, このアルゴリズムが有限回の反復で問題 P の近似最適解を見出すことを示している。

【定理 3】

部分問題の係数行列 $\{H_k\}$ は一様に正定値かつ有界であるとする。そのとき, 生成される点列 $\{x^k\}, \{x^k + p^j\}$, および $\{v^j\}$ が有界ならば, このアルゴリズムは有限回

の反復で停止し, このとき得られている解 x^k は, 問題 $\hat{P}$ に対する次の ε -最適性条件⁶⁾ を満足する。

$$0 \in \partial_{\varepsilon_0} f(x^k) + \nabla F(x^k) + G(x^k)^T v + \varepsilon_1 S^n$$

$$0 \in g(x^k) + \varepsilon_2 S^m$$

ただし v はラグランジュ乗数のベクトル, S^n, S^m はそれぞれ R^n, R^m の閉単位球, $\partial_\varepsilon f(x)$ は f の ε -劣微分⁶⁾ である。また ε_i , $i = 0, 1, 2$, は, ある正定数 κ と H_k の最小, 最大固有値 $\lambda_{\min}, \lambda_{\max}$ を用いて, 次式で与えられる。

$$\varepsilon_0 = \gamma + \delta$$

$$\varepsilon_1 = \left(\frac{\kappa}{\gamma + \delta} + \lambda_{\max} \right) \sqrt{\frac{2\delta}{\lambda_{\min}}} + \sqrt{2(\gamma + \delta)\lambda_{\max}}$$

$$\varepsilon_2 = \delta / \rho$$

証明 定理 2 の結果を用いれば参考文献 6) と同様の方法で証明できるので, ここでは省略する。

アルゴリズム SQPD を問題 $\hat{P}$ に対して適用したとき, 得られた解 $\bar{x}$ が P の最適解の適当な近似となっているかどうか, すなわち, 問題 $\hat{P}$ を定義するときに用いた定数 M_i , $i = 1, \dots, l$, が十分大きいかどうかを判定しなければならない。そのためには, 問題 $\hat{P}_i(\bar{x})$, $i = 1, \dots, l$, の最適解 $\bar{y}_i$ と, それに対応するラグランジュ乗数 $\bar{u}_i$ に対して $(A_i^T \bar{x} + B_i^T \bar{y}_i - b_i)^T \bar{u}_i$, $i = 1, \dots, l$, を計算し, もし $(A_i^T \bar{x} + B_i^T \bar{y}_i - b_i)^T \bar{u}_i > 0$ であるような i が存在すれば, M_i の値が小さいと考えられるので, この M_i を大きくして再びアルゴリズムを実行すればよい。なお, 正確なペナルティ関数 $W(x, r)$ に含まれているパラメータ r は, アルゴリズムの Step 2 によって, 自動的に適当な大きさに調整されることに注意しておく。

4. 数値実験

問題 P に対するアルゴリズムを FORTRAN 77 でコード化し, FACOM M 780-30 を用いて倍精度で数値実験を行なった。

非線形関数 $F(x)$, $g_i(x)$, $i = 1, \dots, m$, は 2 次関数とし, すべての関数の係数は乱数を用いて定めた。また, 2 次計画問題と線形計画問題の解法としてはそれぞれレムケ法とシンプレックス法を採用し, アルゴリズムで使用する正定値対称行列 H_k は, まず H_0 として単位行列 I を用い, 以降の各反復では問題 $\hat{P}$ のラグランジュ関数のヘッセ行列を近似するように BFGS 公式を用いて更新した。ただし, 前節にも述べたように, 問題 $\hat{P}$ には微分可能でない関数が含まれているので, BFGS 公式においてその勾配を用いるところは劣勾配を形式的に代用した。

以下の実験では非線形変数の数 $n = 10$, 非線形制約の

Fig. 1 Computational results for test problems with $l=2$: CPU times for solving three examples for each problem size

Fig. 2 Computational results for test problems with $l=5$: CPU times for solving three examples for each problem size

数 $m=10$ とし、特に断わらない限りパラメータの値を次のように設定した。

$$\gamma = 10^{-2}, \delta = 10^{-4}, \rho = 10^{-2}, \zeta = 1.5 \\ \beta = 0.9, \sigma = 0.5, \tau = 0.5$$

まず、線形部分のブロックの数 $l=2$ として、行列 B_i , $i=1, 2$ の大きさがそれぞれ 20×10 , 40×20 , 60×30 , 80×40 , 100×50 の各場合について、異なる問題を 3 題ずつ生成し、アルゴリズム SQPD を適用した。このときの CPU 時間を Fig. 1 に示す。この場合、たとえば行列 B_i が 100×50 ならば、全体として非線形変数の数が 10, 線形変数の数が 100, 非線形制約の数が 10, 線形制約の数が 200 の問題となる。

つぎに、ブロックの数 $l=5$ として、 $l=2$ の場合と全く同様の実験を行なった。Fig. 2 がその結果である。この場合、たとえば行列 B_i が 100×50 ならば、全体として非線形変数の数が 10, 線形変数の数が 250, 非線形制約の数が 10, 線形制約の数が 500 の問題となる。

さらに従来のアルゴリズムとの比較のため、提案したアルゴリズム SQPD と、英国 National Physical Laboratory で開発された NPL ライブラリーの中の乗数法を用いたサブルーチン SALQDR の両方に対して同じ問題 ($l=2$) を与えて実験を行なった。問題は 10×5 , 20×10 の各場合について 3 題ずつ発生させた。一般に

乗数法では精度の高い解を得ることができる。このため、SQPD によって得られる解の精度を乗数法のそれと同程度にするために、パラメータのうち、次の三つのものについて

$$\gamma = 10^{-5}, \delta = 10^{-10}, \rho = 10^{-5}$$

と設定を変更して解いた。実験結果を Fig. 3 に示す。さらに、各アルゴリズムにより得られた解 x とラグランジュ乗数 v の精度を

$$\Delta_1 = \| \nabla f(x) + v^T G(x) \|_\infty$$

$$\Delta_2 = \| g(x) \|_\infty$$

により評価したところ、すべてのテスト問題に対してアルゴリズム SQPD では $\Delta_1 = 10^{-8} \sim 10^{-6}$, $\Delta_2 = 10^{-13} \sim 10^{-9}$, SALQDR は $\Delta_1 = 10^{-8} \sim 10^{-6}$, $\Delta_2 = 10^{-9} \sim 10^{-6}$, とほぼ同程度の精度の解が得られていることが確認された。

以上の実験結果より、問題 P のような構造を持つ問題に対して、アルゴリズム SQPD は、非常に効率的であることがわかる。この実験で解いた最大の問題は、全体として変数の数 260, 制約条件の数 510 であるが、SQPD はこのような大きな非線形計画問題を取り扱えることも確かめられた。

ここで行なった実験では、正定値対称行列 H_k を準

Fig. 3 Comparison of the proposed algorithm (SQPD) with multiplier method (SALQDR)

ニュートン法の BFGS 公式を用いて更新したが, BFGS 公式はもともと微分可能な問題に対する公式であり, 微分不可能な問題に適用することの理論的な正当性は明確ではない. このような微分不可能問題に対する H_k の効率的な更新法を開発することは今後の課題である.

5. おわりに

大規模数理計画問題の解法に関する従来の研究は, 線形な問題を取り扱うものが主流であった^{1), 2)}. しかし, それは現実に現れるほとんどの問題が線形であるからというのではなく, 単に大規模非線形計画問題を効率的に解きうる手法の開発が容易でないからというのがその理由と考えられる. 本論文では, 微分不可能な最適化問題に対する最新の手法を取り入れることにより, ある種の構造をもつ非線形計画問題に対して新しい分解法を提案し, その有効性を数値実験により確認した. また, このような分解法は, 大規模問題に対する手法としてだけでなく, 複数のプロセッサを同時に使用して計算を行なう. いわゆる並列計算の枠組みにおいても今後重要性を増していくものと考えられる.

なお本研究は一部文部省科学研究費補助金 (一般研究 (C) 02680025) によるものである.

参考文献

- 1) L. S. Lasdon : Optimization Theory for Large Systems, Macmillan (1970)
- 2) J. F. Shapiro : Mathematical Programming ; Structures and Algorithms, Wiley (1979)
- 3) A. M. Geoffrion : Generalized Benders Decomposition ; J. Opt. Theory Appl., Vol. 10, pp. 237~260 (1972)
- 4) O. L. Mangasarian : Sufficiency of Exact Penalty Minimization ; SIAM J. Control & Opt., Vol. 23, pp. 30~37 (1985)
- 5) R. Lazimy : An Exact Penalty Function and Relaxation Approach for Solving Decomposable Nonlinear Programs ; SIAM J. Control & Opt., Vol. 26, pp. 127~137 (1988)
- 6) M. Fukushima : A Successive Quadratic Programming Method for a Class of Constrained Nonsmooth Optimization Problems ; Math. Programming, Vol. 49, pp. 231~251 (1991)
- 7) A. Auslender : Numerical Methods for Nondifferentiable Convex Optimization ; Math. Programming Study, Vol. 30, pp. 102~126 (1987)
- 8) 福島 : 非線形最適化の理論, 産業図書 (1980)

付 録

付録 1 定理 1 の証明

問題 $SP(x)$ の Kuhn-Tucker 条件は次のようになる.

$$\begin{aligned} 0 &\in \partial f(x + \bar{p}) + \nabla F(x) + H\bar{p} + G(\bar{x})^T \bar{v} \\ g(x) + G(x)\bar{p} &\leq 0, \quad \bar{v} \geq 0 \\ \bar{v}^T \{g(x) + G(x)\bar{p}\} &= 0 \end{aligned}$$

これと問題 $\hat{P}$ の Kuhn-Tucker 条件の比較より, $\bar{p} = 0$ のとき $(x, \bar{v})$ が問題 $\hat{P}$ の Kuhn-Tucker 条件を満足することは明らかである.

つぎに $\bar{p} \neq 0$ の場合を考える. まず任意の x, d に対して W の方向微係数 W'_d を次で定義する.

$$W'_d(x, r) = \lim_{\alpha \rightarrow 0} \frac{W(x + \alpha d, r) - W(x, r)}{\alpha}$$

さらに添字集合 I_1, I_2, I_3 を次のように定める.

$$\begin{aligned} I_1 &= \{i | g_i(x) > 0\} \\ I_2 &= \{i | g_i(x) = 0\} \\ I_3 &= \{i | g_i(x) < 0\} \end{aligned}$$

ここで, 問題 $SP(x)$ の Kuhn-Tucker 条件より

$$\bar{\xi} + \nabla F(x) + H\bar{p} + G(x)^T \bar{v} = 0$$

をみたす $\bar{\xi} \in \partial f(x + \bar{p})$ が存在する. また f は凸関数であるから,

$$\max\{\xi^T \bar{p} | \xi \in \partial f(x)\} \leq \bar{\xi}^T \bar{p}$$

このとき, $\bar{v} = (\bar{v}_1, \dots, \bar{v}_m)$ として

$$\begin{aligned} W'_p(x, r) \\ = \max\{\xi^T \bar{p} | \xi \in \partial f(x)\} + \nabla F(x)^T \bar{p} \end{aligned}$$

$$\begin{aligned}
& +r \sum_{i \in I_1} \nabla g_i(x)^T \bar{p} + r \sum_{i \in I_2} \max\{\nabla g_i(x)^T \bar{p}, 0\} \\
& \leq \max\{\xi^T \bar{p} \mid \xi \in \partial f(x)\} - \xi^T \bar{p} - \bar{p}^T H \bar{p} \\
& \quad - \sum_{i \in I_1} (r - \bar{v}_i) g_i(x) + \sum_{i \in I_2} \bar{v}_i g_i(x)
\end{aligned}$$

H は正定値であるから、 $r \geq \|\bar{v}\|_\infty$ ならば

$$W_{\bar{p}}(x, r) < 0$$

が成り立つ。(証明おわり)

付録2 定理2の証明

問題 $QP_j(x)$ の Kuhn-Tucker 条件は次のように書ける。

$$\begin{aligned}
& \lambda_j \nabla h_{j-1}(p^j) + (1 - \lambda_j) \nabla h_{-(j-1)}(p^j) + \nabla F(x) \\
& + H p + G(x)^T v^j = 0 \\
& g(x) + G(x) p^j \leq 0, \quad v^j \geq 0 \\
& v^{jT} \{g(x) + G(x) p^j\} = 0
\end{aligned}$$

このとき次の問題 $\widehat{QP}_j(x)$ について考える。

$$\begin{aligned}
\widehat{QP}_j(x): & \text{minimize } \hat{\Psi}_j(p) \\
& \text{subject to } g(x) + G(x) p \leq 0
\end{aligned}$$

ただし、関数 $\hat{\Psi}_j: R^n \rightarrow R$ は次で定義する。

$$\hat{\Psi}_j(p) = h_{-j}(p) + F(x) + \nabla F(x)^T p + \frac{1}{2} p^T H p$$

$h_{-j}(p) = \lambda_j h_{j-1}(p) + (1 - \lambda_j) h_{-(j-1)}(p)$ であるから、 p^j, v^j は問題 $\widehat{QP}_j(x)$ の Kuhn-Tucker 条件も満たす。したがって、問題 $\widehat{QP}_j(x)$ が凸計画問題であることを考えると、 $g(x) + G(x) p \leq 0$ を満たす p に対して

$$\Psi_j(p^j) = \hat{\Psi}_j(p^j) \leq \hat{\Psi}_j(p)$$

が成立する。ここで定義より、任意の p に対して $\hat{\Psi}_j(p) \leq \Psi_{j+1}(p)$ であるから、

$$\Psi_j(p^j) \leq \Psi_{j+1}(p^{j+1})$$

さらに劣微分の定義⁸⁾により $h_j(p) \leq f(x+p)$, $j=1, 2, \dots$, が成立し、また $0 \leq \lambda_j \leq 1$ であるので $h_{-j}(p) \leq f(x+p)$ も成り立つ。したがって $\Psi_j(p) \leq \Phi(p)$ である。これにより、 $j=1, 2, \dots$, に対して

$$\Psi_j(p^j) \leq \Psi_{j+1}(p^{j+1}) \leq \Psi_{j+1}(\bar{p}) \leq \Phi(\bar{p})$$

が成り立つ。以上より (1) が証明された。

つぎに (2)~(4) を証明する。まず仮定より $\{p^j\}, \{v^j\}$ は有界であり、またそれによって $\{\xi^j\}$ も有界である⁹⁾。ここで $\hat{\Psi}_j$ の強凸性⁹⁾により、 H の最小固有値 $\lambda_{\min}$ に対して

$$\hat{\Psi}_j(p^j) + \frac{\lambda_{\min}}{2} \|p^{j+1} - p^j\|^2 \leq \hat{\Psi}_j(p^{j+1})$$

が成立する。このとき

$$\Psi_j(p^j) = \hat{\Psi}_j(p^j) \leq \hat{\Psi}_j(p^{j+1}) \leq \Psi_{j+1}(p^{j+1}) \leq \Phi(\bar{p})$$

であるから $\hat{\Psi}_j(p^{j+1}) - \hat{\Psi}_j(p^j) \rightarrow 0$ である。したがって

$$\|p^{j+1} - p^j\| \rightarrow 0$$

一方、問題 QP_j の最適解 p^j において $h_{-j}(p^j) \leq f(x+p^j) = h_j(p^j)$ であるから、

$$\begin{aligned}
\Psi_{j+1}(p^{j+1}) &= h_j(p^{j+1}) + F(x) + \nabla F(x)^T p^{j+1} \\
&\quad + \frac{1}{2} (p^{j+1})^T H p^{j+1}
\end{aligned}$$

と書くことができる。また、定義より

$$h_j(p) = f(x+p) + (\xi^j)^T (p - p^j)$$

$$\Phi(p) = f(x+p) + F(x) + \nabla F(x)^T p + \frac{1}{2} p^T H p$$

であるから、

$$\begin{aligned}
\Psi_{j+1}(p^{j+1}) - \Phi(p^j) &= \{\xi^j + \nabla F(x) + \frac{1}{2} H(p^{j+1} + p^j)\}^T (p^{j+1} - p^j)
\end{aligned}$$

が成り立つ。したがって、 $\{p^j\}, \{\xi^j\}$ の有界性より

$$\Psi_{j+1}(p^{j+1}) - \Phi(p^j) \rightarrow 0$$

ここで、 $\Psi_{j+1}(p^{j+1}) \leq \Phi(\bar{p}) \leq \Phi(p^j)$ であることを考えると

$$\Psi_j(p^j) \rightarrow \Phi(\bar{p})$$

$$\Phi(p^j) \rightarrow \Phi(\bar{p})$$

最後に (5) 式を証明する。それには $\{v^j\}$ の任意の集積点 $\bar{v}$ に対して $\bar{v} \in \Lambda$ であることを示せばよい。まず問題 $SP(x)$ に対する一次独立制約想定より、 Λ は有界であることに注意する。つぎに

$$S(t) = \min\{\Phi(p) \mid g(x) + G(x) p \leq t\}$$

$$S_j(t) = \min\{\Psi_j(p) \mid g(x) + G(x) p \leq t\}$$

を定義すると、 $S(t), S_j(t)$ は凸関数であり、 $-v^j \in \partial S_j(0)$ を満たす¹⁰⁾ので、任意の $t \in R^m$ に対して

$$S(t) \geq S_j(t) \geq S_j(0) - (v^j)^T t$$

が成り立つ。ここで $j \rightarrow \infty$ とすれば、(2) により $S_j(0) \rightarrow S(0)$ であるから、

$$S(t) \geq S(0) - \bar{v}^T t$$

が任意の t に対して成立する。すなわち $-\bar{v} \in \partial S(0)$ である。 $\Lambda = -\partial S(0)$ であるから、 $\bar{v} \in \Lambda$ が示された。

(証明おわり)