

# 共役勾配法

矢部 博 東京理科大学

本稿では、無制約最小化問題に対する有力な数値解法の1つである共役勾配法について説明する。

## 1. 共役性とは

### 2次関数

$$f(x) = (1/2)x^T A x - b^T x + c \quad (1)$$

を最小化することを考えよう。ただし、 $x \in \mathbb{R}^n$ は変数ベクトル、 $A \in \mathbb{R}^{n \times n}$ は正定値対称行列、 $b \in \mathbb{R}^n$ は定ベクトル、 $c$ は定数であり、“ $T$ ”は転置を表わす。

この関数を最小化する最も素朴な数値解法は、順次、各座標軸方向にそって目的関数の最小化をくりかえす、いわゆる緩和法であろう。緩和法はしかし、座標軸に対する目的関数の位置によって、その効率がいちじるしく異なる。2変数の場合を例にとれば、目的関数の等高線の楕円が座標軸に対して、ななめに細く横たわる場合に

は図1(a)のように、なかなか最小点に近づかないのだが、一方において楕円の軸が座標軸と平行な場合や、もともと等高線が同心円の場合には、わずか2回( $n$ 変数の場合なら $n$ 回)の直線最小化で最小点に到達してしまう(図1(b))。しかし、このことは逆に緩和法の考え方が、2次関数最小化のための合理的な方法を示唆していることを意味する。すなわち行列を $A = P^T P$  ( $P \in \mathbb{R}^{n \times n}$ )と分解して、 $\hat{x} = Px$ と変数変換すれば、 $\hat{x}$ の空間では(1)の等高線は同心円になるから、緩和法によれば高々 $n$ 回の直線最小化で最小点に到達することができる。このとき1次変換された空間での直交性  $(Pu)^T (Pv) = 0$  は元の空間で

$$u^T A v = 0 \quad (2)$$

を意味するから、結局(2)を満たす零でないベクトルに沿って順々に2次関数の最小化をくりかえしてゆけば、高々 $n$ 回の探索で最小点を得られることが期待される。

さて、零でないベクトル $u, v$ が(2)を満たすとき $u$ と $v$ は( $A$ に関して)互いに共役である、あるいは(2)の意味で直交するという(図2)。

共役方向の作り方にはいろいろあるが、本稿では2次関数(1)の勾配ベクトル

$$\nabla f(x) = Ax - b \quad (3)$$

を利用した共役勾配法(Conjugate Gradient法、以下CG法と呼ぶ[1,2])を紹介する。CG法は連立1次方程式( $Ax = b$ )の数値解法としてHestenes-Stiefel(1952)が提案したもので、Fletcher-Reeves(1964)によって一般の非線形最小化問題への適用が試みられた。その後、理論面と実用面の両面で活発な研

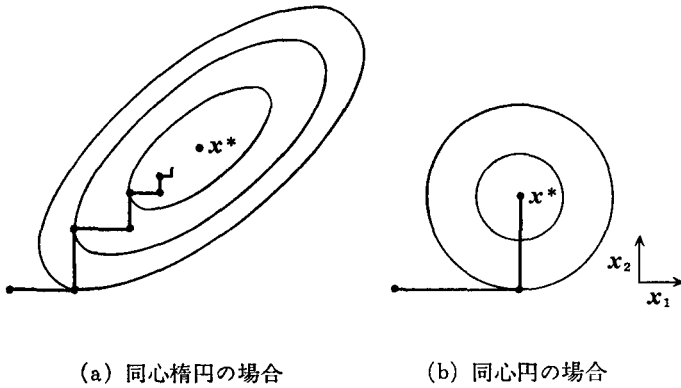
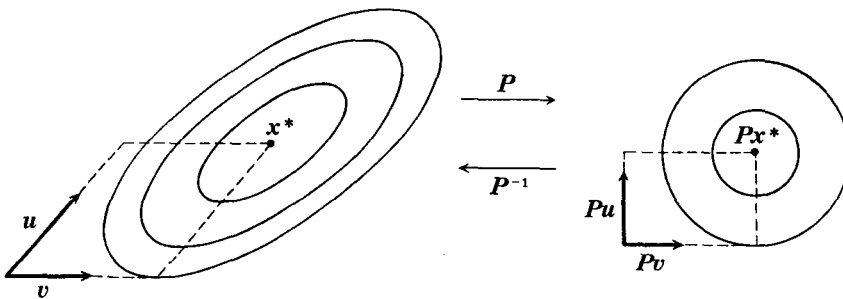


図1 緩和法による2次関数の最小化



(a)  $u$ と $v$ は $A$ に関して互いに共役である (b)  $Pu$ と $Pv$ は互いに直交する

図2 共役性と直交性の関係

究が行なわれて今日に至っている。特に近年では、たとえば原子炉の安全設計や VLSI の開発などで扱われる非常に大型な連立 1 次方程式に対する数値解法として CG 法の系統が利用され、さらにスーパーコンピュータの普及ともあいまってますます研究されている [4, 5].

## 2. 共役方向の生成

では、どのようにして行列 A に関して互いに共役な方向ベクトルを生成すればよいのであろうか。以下ではベクトル  $u$  と  $v$  の内積として

$$(u, v)_A = u^T A v \quad (4)$$

を導入し、内積 (4) の意味で互いに直交するベクトルを Gram-Schmidt の方法で生成することを考えよう。

まず、1 次独立な 2 つのベクトル  $v_0$  と  $v_1$  が与えられている場合を考える。このとき  $v_1$  を  $v_0$  へ (4) の意味で正射影して得られるベクトルは  $\gamma_{10} v_0$  と表わされるから、 $d_1 = v_1 - \gamma_{10} v_0$  が  $(d_1, v_0)_A = 0$  を満たすように係数  $\gamma_{10}$  を決めれば

$$d_1 = v_1 - \{(v_0, v_1)_A / (v_0, v_0)_A\} v_0 \quad (5)$$

が得られる (図 3)。よって  $v_0$  と  $v_1$  を用いて、(4) の意味で直交するベクトル  $d_0 = v_0$  と  $d_1$  が生成された。ここで、 $v_0$  と  $v_1$  が張る部分空間を  $\text{span}[v_0, v_1]$  で表わせば、 $\text{span}[v_0, v_1] = \text{span}[d_0, d_1]$  が成り立つことに注意しておく。

次に  $\text{span}[d_0, d_1]$  に対して 1 次独立なベクトル  $v_2$  が与えられた場合を考える。このとき  $v_2$  を  $d_0, d_1$  へ (4) の意味で正射影して得られるベクトルはそれぞれ

$$\gamma_{20} d_0, \quad \gamma_{20} = (d_0, v_2)_A / (d_0, d_0)_A,$$

$$\gamma_{21} d_1, \quad \gamma_{21} = (d_1, v_2)_A / (d_1, d_1)_A$$

で表わされるから、 $\gamma_{20} d_0 + \gamma_{21} d_1$  が  $v_2$  の  $\text{span}[d_0, d_1]$  上

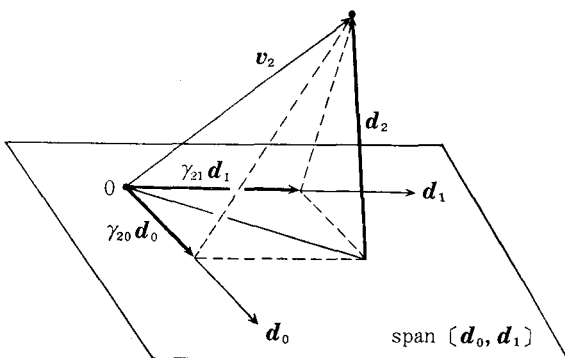


図 4  $d_0, d_1$  に共役なベクトル  $d_2$  を生成する ( $R^3$  の場合)

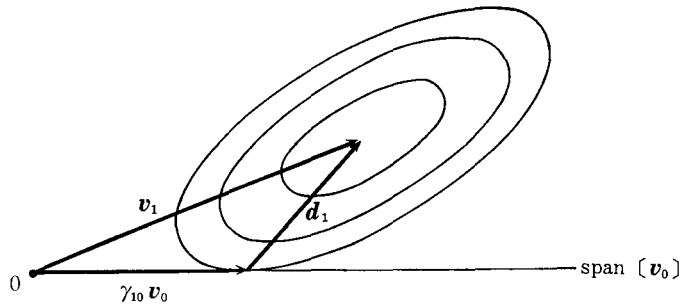


図 3  $v_0$  と  $v_1$  から共役方向を生成する ( $R^2$  の場合)

への (4) の意味での正射影になっている。したがって

$$d_2 = v_2 - \{(d_0, v_2)_A / (d_0, d_0)_A\} d_0 - \{(d_1, v_2)_A / (d_1, d_1)_A\} d_1 \quad (6)$$

が (4) の意味で  $d_0$  と  $d_1$  に直交するベクトルになる (図 4)。

以上の手順を一般化すれば、順々に与えられるベクトル  $v_k$  に対して

$$d_0 = v_0, \\ d_k = v_k - \sum_{i=0}^{k-1} \{(d_i, v_k)_A / (d_i, d_i)_A\} d_i, \quad k \geq 1 \quad (7)$$

によって共役方向が生成される。さらに図 3, 4 から明らかのように、各  $k$  に対して

$$\text{span}[v_0, \dots, v_k] = \text{span}[d_0, \dots, d_k] \quad (8)$$

が成り立つ。

## 3. 2 次関数最小化に対する CG 法

前節で導いた共役方向の生成手順 (7) の中で、毎回追加されていく 1 次独立なベクトル  $v_k$  の選び方には自由度があるが、標準の (Hestenes-Stiefel 流の) CG 法では、与えられた初期点  $x_0$  での最急降下方向  $-\nabla f(x_0)$  が  $v_0$  として選ばれる。そして  $k$  回目の反復では、 $d_k$  方向で 2 次関数を最小にする刻み幅  $\alpha_k$  を選んで (この作業を「正確な直線探索」というが、以下では単に「直線探索」と呼ぶことにする)、次の点を

$$x_{k+1} = x_k + \alpha_k d_k \quad (9)$$

とし、その点での最急降下方向  $-\nabla f(x_{k+1})$  を  $v_{k+1}$  に選んでいく。そのさい、共役方向は (7) で生成されていくわけであるが、各  $k$  に対して

$$\nabla f(x_{k+1})^T \nabla f(x_j) = 0, \quad j = 0, \dots, k \quad (10)$$

が成り立つことに注意すれば (7) の和の部分は非常に簡単になり、新しい探索方向は

$$d_{k+1} = -\nabla f(x_{k+1}) + \beta_k d_k, \\ \beta_k = d_k^T A \nabla f(x_{k+1}) / d_k^T A d_k \quad (11)$$

で与えられる。ここで特徴的なことは、新しい探索方向

$\mathbf{d}_{k+1}$  がその点での最急降下方向  $-\nabla f(\mathbf{x}_{k+1})$  と前回の探索方向  $\mathbf{d}_k$  の2本だけから生成されていくことである。これは初回で、 $\mathbf{v}_0 = -\nabla f(\mathbf{x}_0)$  とおいたことが本質的に効いている。

また、 $\beta_k$  は2次関数に対していろいろと変形できていくつかの変種が考えられており、特に次の2つの公式が有名である。

(i) Fletcher-Reeves の公式

$$\beta_k = \frac{\nabla f(\mathbf{x}_{k+1})^T \nabla f(\mathbf{x}_{k+1})}{\nabla f(\mathbf{x}_k)^T \nabla f(\mathbf{x}_k)}, \quad (12)$$

(ii) Polak-Ribière-Polyak の公式

$$\beta_k = \frac{\nabla f(\mathbf{x}_{k+1})^T (\nabla f(\mathbf{x}_{k+1}) - \nabla f(\mathbf{x}_k))}{\nabla f(\mathbf{x}_k)^T \nabla f(\mathbf{x}_k)} \quad (13)$$

ただし一般の目的関数にCG法を適用する場合には、 $\beta_k$  の選び方によって計算効率がかかなり変わることがある[3]。

CG法の本質的な性質は  $\nabla f(\mathbf{x}_{k+1})^T \mathbf{d}_j = 0 (j=0, \dots, k)$  が成り立つことである。すなわち順々に得られていく点  $\mathbf{x}_{k+1}$  はアフィン空間  $\mathbf{x}_0 + \text{span}[\mathbf{d}_0, \dots, \mathbf{d}_k]$  上の  $f$  の最小点になる[3] (図5)。

さらに式(8)と同様に  $\text{span}[\mathbf{d}_0, \dots, \mathbf{d}_k] = \text{span}[\nabla f(\mathbf{x}_0), \dots, \nabla f(\mathbf{x}_k)] = \text{span}[\nabla f(\mathbf{x}_0), A \nabla f(\mathbf{x}_0), \dots, A^k \nabla f(\mathbf{x}_0)]$  が成り立つことも重要である。またCG法では、式(9)によって点列  $\{\mathbf{x}_k\}$  が生成されていくわけであるが、図5の見方を変えれば、アフィン空間  $\mathbf{x}_0 + \text{span}[\nabla f(\mathbf{x}_0), \dots, A^k \nabla f(\mathbf{x}_0)]$  へ内積(4)の意味で最小点  $\mathbf{x}^*$  を正射影することによって次の点  $\mathbf{x}_{k+1}$  が生成されるとも解釈できる[6] (図6)。

図5と図6のいずれの解釈にせよ、高々  $n$  回の手順で

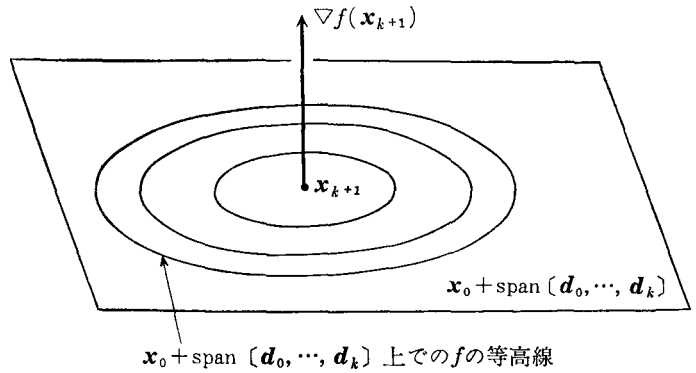


図5  $\mathbf{x}_{k+1}$  はアフィン空間  $\mathbf{x}_0 + \text{span}[\mathbf{d}_0, \dots, \mathbf{d}_k]$  上の最小点である

最小点を得られることが示され、特に行列Aの固有値が重複していればいほど、もっと早い時期に解が得られる[7]。こうした性質をもつことを考慮すれば、CG法は直接法と反復法の折衷法ともみなせるであろう。

本節を終えるにあたって、CG法がどのようにして点列と探索方向を生成していくかを、3次元空間でながめてみよう。初期点  $\mathbf{x}_0$  が与えられたとき、初回の探索方向は  $\mathbf{d}_0 = -\nabla f(\mathbf{x}_0)$  である。ただし、 $-\nabla f(\mathbf{x}_0)$  は曲面  $f(\mathbf{x}) = f(\mathbf{x}_0)$  に対する点  $\mathbf{x}_0$  での接平面の法線ベクトルで、 $f$  の最急降下方向である(図7)。ここで  $\mathbf{d}_0$  方向での直線探索を行なえば、 $\mathbf{x}_1$  が得られる。このときベクトル  $\mathbf{d}_0$  を延長した半直線は点  $\mathbf{x}_1$  で曲面  $f(\mathbf{x}) = f(\mathbf{x}_1)$  に接するから、 $\mathbf{d}_0^T \nabla f(\mathbf{x}_1) = 0$  が成り立つ。よって、 $\mathbf{d}_0$  の選び方より  $\nabla f(\mathbf{x}_0)^T \nabla f(\mathbf{x}_1) = 0$  がいえる。

次に2回目の反復を考えよう。初期点および2本のベクトル  $\mathbf{d}_0, \nabla f(\mathbf{x}_1)$  とで作られるアフィン空間  $\mathbf{x}_0 + \text{span}[\mathbf{d}_0, \nabla f(\mathbf{x}_1)]$  で目的関数の等高曲面を切れば、その切り口は図8に示すような等高線になる。このとき、 $\mathbf{d}_0$  方向に適当に長さを調節したベクトル  $\beta_0 \mathbf{d}_0$  と、点  $\mathbf{x}_1$

での最急降下方向  $-\nabla f(\mathbf{x}_1)$  との和から作られる新しい探索方向  $\mathbf{d}_1$  は、このアフィン空間上での  $f$  の最小点  $\mathbf{x}_2$  を向いているので、 $\mathbf{d}_1$  方向で直線探索を行なえば点  $\mathbf{x}_2$  が得られる。図から明らかのように  $\text{span}[\mathbf{d}_0, \nabla f(\mathbf{x}_1)] = \text{span}[\mathbf{d}_0, \mathbf{d}_1]$  が成り立ち、また  $\nabla f(\mathbf{x}_2)$  は部分空間  $\text{span}[\mathbf{d}_0, \nabla f(\mathbf{x}_1)]$  と直交するから

$$\begin{aligned} \nabla f(\mathbf{x}_2)^T \nabla f(\mathbf{x}_0) &= \nabla f(\mathbf{x}_2)^T \\ \nabla f(\mathbf{x}_1) &= 0, \end{aligned}$$

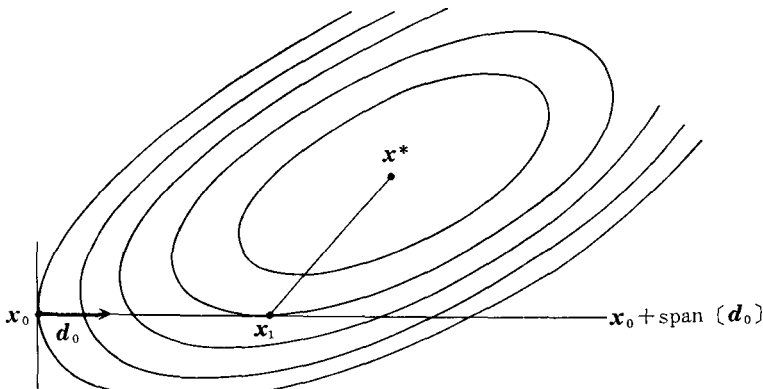


図6  $\mathbf{x}^*$  の  $\mathbf{x}_0 + \text{span}[\mathbf{d}_0]$  上への射影として  $\mathbf{x}_1$  を得る ( $\mathbf{R}^2$  の場合)

$$\nabla f(\mathbf{x}_2)^T \mathbf{d}_0 = \nabla f(\mathbf{x}_2)^T \mathbf{d}_1 = 0$$

が成り立つ (図 8).

さらに 3 回目の反復では,  $\mathbf{d}_1$  方向に適当に長さを調節したベクトル  $\beta_1 \mathbf{d}_1$  と, 点  $\mathbf{x}_2$  での最急降下方向  $-\nabla f(\mathbf{x}_2)$  との和から作られる新しい探索方向  $\mathbf{d}_2$  は, アフィン空間  $\mathbf{x}_0 + \text{span}[\mathbf{d}_0, \mathbf{d}_1, \nabla f(\mathbf{x}_2)]$  上での  $f$  の最小点, すなわち全空間での  $f$  の最小点  $\mathbf{x}^*$  を向いている. したがって,  $\mathbf{d}_2$  方向の直線探索によって解が得られるから, CG法が終了する (図 9).

#### 4. 大型の連立 1 次方程式 に対する CG 法

行列  $A$  が正定値対称ならば, 2 次関数 (1) を最小化することと連立 1 次方程式  $A\mathbf{x}=\mathbf{b}$  を解くことが同等になるので, ここでは大型の連立 1 次方程式に実際に CG法を適用する場合の工夫について少し触れておこう. ただし実際の数値計算では丸め誤差の影響などで必ずしも  $n$  回の手順で解に到達できるとは限らない.

行列  $A$  の固有値に重複があるときや密集固有値があるときには, CG法は  $n$  回より少ない探索回数で十分精度のよい近似解を見つけることができる. そこで  $A\mathbf{x}=\mathbf{b}$  をまともに解く代わりに, 何らかの前処理を行なって, 係数行列が密集固有値をもつような問題に変換しておいてから CG法を適用すれば, 効率が上がることが期待される. たとえば  $A \approx LL^T$  ( $L$  は適当な下三角行列) と不完全コレスキー (Incomplete Cholesky) 分解して

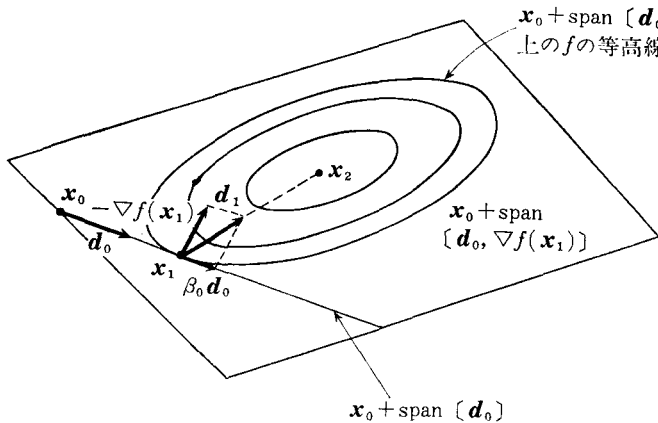


図 8 CG法の 2 回目の反復

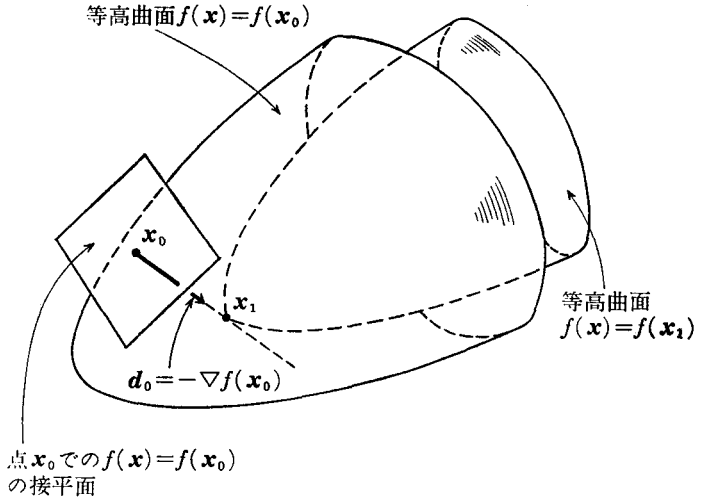


図 7 CG法の 1 回目の反復

$$(L^{-1}AL^{-T})(L^T\mathbf{x})=L^{-1}\mathbf{b} \quad (14)$$

なる連立 1 次方程式を解くことを考えてみよう, このとき係数行列  $L^{-1}AL^{-T}$  の固有値はかなり密集 (正確にコレスキー分解すれば固有値はすべて 1) しているだろうから, CG法を適用すれば少ない探索回数で精度の良い近似解が得られるであろう. この方法は頭文字をとって **ICCG 法** と呼ばれている [4]. このような前処理 (Preconditioning) を組み込んだ CG法を総称して **PCG 法** と呼んでおり, 近年ではスーパーコンピュータに PCG 法をいかにうまく組み込んで計算効率を上げるかが盛んに検討されている [5].

さらに係数行列  $A$  の正定値性が崩れた方程式や, 行列  $A$  が対称ですらないような方程式にも CG法の系統が適用され, **BCG (Biconjugate Gradient) 法** や **CR (Conjugate Residual) 法** などが考案されている [4, 6]. こうした解法は単なる理論にとどまらず, 今後ますます実用化されていくことであろう.

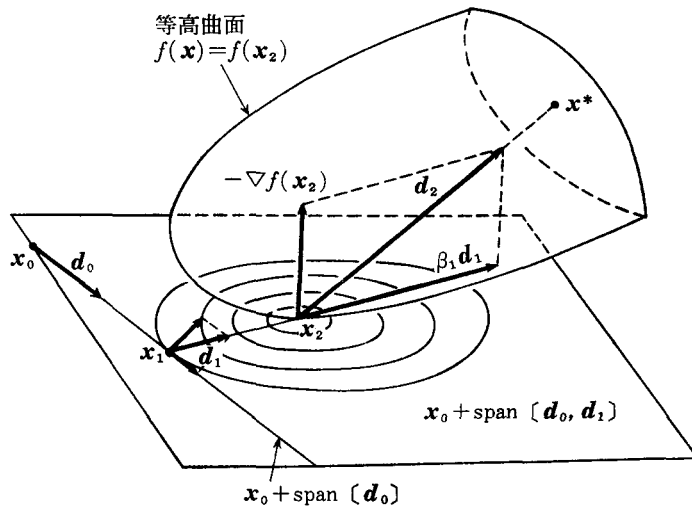


図 9 CG法の3回目の反復

参 考 文 献

[1] M. R. Hestenes: HISTORICAL PAPER Iterative methods for solving linear equations, *Journal of Optimization Theory and Applications*, Vol. 11, pp. 323-334(1973).

[2] M. R. Hestenes: *Conjugate Direction Methods in Optimization*, Springer-Verlag, Berlin (1980).

[3] 今野浩, 山下浩: 非線形計画法, ORライブラリー 6, 日科技連 (1978).

[4] 村田健郎, 小国力, 唐木幸比古: スーパーコンピュータ, 丸善 (1985).

[5] M. Natori and T. Nodera Eds.: *Proceedings*

of Symposium on Large Sparse Sets of Linear Equations, No.1 ("Conjugate Gradient Method and its Applications", 1986), No.2 ("Conjugate Gradient Method and Supercomputer", 1986), Keio University.

[6] J. Stoer: Solution of large linear systems of equations by conjugate gradient type methods, in *Mathematical Programming—The State of the Art* (A. Bachem, M. Grötschel and B. Korte Eds.), pp. 540-565, Springer-Verlag, Berlin (1983).

[7] 戸川隼人: 共役勾配法, 新しい応用の数学17, 教育出版 (1977).

次 号 予 告

特集 交通

宅急便	齊木 勝彦 (ヤマト運輸)
消費者主導の物流新時代	忍田 和良 (日通総研)
東京ディズニー・ランド建設とレジャー産業の本質	丹沢 章治 (三井不動産)
東京圏における高速鉄道を中心とする交通網の整備に関する基本計画について	北野 嘉幸 (JR西日本)
総合交通システムとインターフェース	上田亀之助 (上田イノベーション研究所)
浮上式鉄道—時速400キロを達成	田中 寿 (鉄道総研)
旅行計画立案システム (TPSS) について	矢部 眞, 八戸英夫 (工学院大)