A Note on the Conjugate Gradient Method

IWARŌ TAKAHASHI*

1. *Preface*

The *conjugate gradient method* which solves simultaneous linear equations $Ax=b$ is well known by numerical analysts. This paper has two purposes. One is to generalize the gradient method which searches the maximum point of a concave function $f(x)=f(x_1, \ldots, x_N)$. The other is to show that our *generalized gradient method,* applied to solve simultaneous linear equations, is itself the conjugate gradient method.

2. *Gradient Method*

The *gradient method* is the recurrence formula

$$x^{n+1}=x^n+h^n\partial f(x^n), \tag{1}$$

where $\partial f=(\partial f/\partial x_1, \ldots, \partial f/\partial x_N)$, with an arbitrary initial approximation x^0.

The h^n, which is a positive scalor, is determined to maximize

$$f(x^n+h^n\partial f(x^n)). \tag{2}$$

If $f(x)$ is differentiable** and concave**, maximizing point of (2) coincides with the stationary point, at which $f(x)$ has zero derivatives.

Above procedures determine x^{n+1} only with the informations about x^n. We could say, this relation is *simple Markoff type.* If x^{n+1} is to be determined by $f(x^n)$ which is the information at x^n, and the direction p^{n-1} which has been used to determine x^n, then we get a new *generalized gradient method.*

3. *Generalized Gradient Method*

(i) *Initial Step*

Select an arbitrary initial approximation x^0, and compute

$$p^0=\partial f(x^0) \tag{3}$$

$$x^1=x^0+\alpha^0 p^0 \tag{4}$$

where $\alpha^0>0$ is determined to maximize

$$f(x^0+\alpha^0 p^0), \tag{5}$$

that is,

$$[\partial f(x^0+\alpha^0 p^0),\ p^0]=0^{***} \tag{6}$$

This paper first appeared in Japanese in Joho Shori (the Journal of the Information Processing Society of Japan), Vol. 5, No. 4 (1964), pp. 203~205.

* Institute for Research in Productivity, Waseda University, Tokyo.

** Hereafter we assume $f(x)$ has these properties.

*** $[A, B]$ denotes inner-product of vectors A and B.

(ii) *Iterative Step* $(n=1, 2, \ldots)$

$$p^n=\partial f(x^n)+\beta^n p^{n-1} \tag{7}$$

$$x^{n+1}=x^n+\alpha^n p^n, \tag{8}$$

where scalors α^n and β^n are determined to maximize

$$f(x^{n+1})=f(x^n+\alpha^n(\partial f(x^n)+\beta^n p^{n-1}))$$

that is, by the conditions

$$[\partial f(x^{n+1}),\ p^n]=0 \tag{9}$$

$$[\partial f(x^{n+1}),\ p^{n-1}]=0 \tag{9'}$$

Above procedures produce $x^1, x^2, x^3, \ldots$ converging the maximum point of $f(x)$.

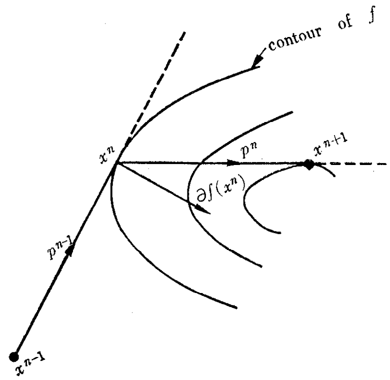

Fig. 1.

4. *Equivalence of Simultaneous Linear Equations and Maximizing Problem of a Quadratic Function*

(i) *The Case with a Symmetric and Positive Definite Coefficient Matrix*

If a matrix A is symmetric and positive definite, then to solve simultaneous linear equations

$$Ax=b \tag{10}$$

is equivalent to maximizing

$$f(x)=2b'x-x'Ax. \tag{11}$$

Because of positive definiteness of A, A^{-1} exists, and we have, for all x,

$$(x-A^{-1}b)'A(x-A^{-1}b)\geqq 0. \tag{12}$$

If we have $=$ in (12) instead of $\geqq$, then

$$x-A^{-1}b=0. \qquad (13)^*$$

(ii) *General Case*

If B is a non-singular matrix, to solve

* If A is symmetric and positive definite, $y'Ay=0$ implies $y=0$.

$$Bx=d \tag{15}$$

is equivalent to minimize

$$(d-Bx)'(d-Bx)=d'd-2d'Bx+x'B'Bx. \tag{16}$$

From (15) we have

$$B'Bx=B'd. \tag{17}$$

If B is non-singular $B'B$ is symmetric and positive definite. So (17) belongs to case (i).

Hereafter we will consider only case (i) with (10).

5. *Generalized Gradient Method Applied to Solve Simultaneous Linear Equations*

Now we will apply our generalized gradient method to $f(x)$ in (11). We have

$$\partial f(x)/2=b-Ax, \tag{19}$$

which we call residual vector r. Let us set

$$r^n=b-Ax^n \quad (=\partial f(x^n)/2) \quad (n=0, 1, 2, \ldots.), \tag{20}$$

then from (8) we have

$$r^{n+1}=r^n-\alpha^n Ap^n \quad (n=0, 1, 2, \ldots.). \tag{21}$$

Now, α^n and β^n are determined as follows. From (6) we have

$$[r^1, p^0]=[r^0-\alpha^0 Ap^0, p^0], \tag{22}$$

so

$$\alpha^0=[r^0, p^0]/[p^0, Ap^0]. \tag{23}$$

From (9′) we have

$$[r^{n+1}, p^{n-1}]=0 \quad (n=1, 2, \ldots.), \tag{24}$$

which with (21) becomes

$$[r^n-\alpha^n Ap^n, p^{n-1}]=0 \quad (n=1, 2, \ldots.). \tag{25}$$

From (9) we have

$$[r^{n+1}, p^n]=0 \quad (n=1, 2, \ldots.), \tag{26}$$

which with (21) becomes

$$[r^n-\alpha^n Ap^n, p^n]=0 \quad (n=0, 1, 2, \ldots.), \tag{28}$$

so

$$\alpha^n=[r^n, p^n]/[p^n, Ap^n] \quad (n=1, 2, \ldots.). \tag{30}$$

From (22), (25) and (26) we have

$$[Ap^n, p^{n-1}]=0 \quad (n=1, 2, \ldots.), \tag{27}$$

which with (7) and (19) becomes

$$[A(2r^n+\beta^n p^{n-1}), p^{n-1}]=0 \quad (n=1, 2, \ldots.), \tag{29}$$

so

$$\beta^n=-2[p^{n-1}, Ar^n]/[p^{n-1}, Ap^{n-1}] \quad (n=1, 2, \ldots.). \tag{31}$$

(Note: In (23) because of positive definiteness of $A[p^0, Ap^1]\neq 0$ so long as

$r^0=p^0\neq 0$. In (30) and (31) $p^{n-1}\neq 0$, from (26) (7) and (19), so long as $r^{n-1}\neq 0$, therefore $[p^{n-1}, Ap^{n-1}]\neq 0$.)

Let us summarize the procedures to solve $Ax=b$ with our generalized gradient method:

Select arbitrarily initial vector x^0 and let us set

$$p^0=r^0=b-Ax^0, \tag{32}$$

and iterate

$$x^{n+1}=x^n+\alpha^n p^n, \tag{33}$$

$$r^{n+1}=r^n-\alpha^n Ap^n, \tag{34}$$

$$p^{n+1}=2r^{n+1}+\beta^{n+1}p^n \quad (n=0, 1, 2, \ldots.), \tag{35}$$

with α^n, β^n determined from (23) (30) and (31).

This algorithm is the conjugate gradient method itself.

6. *Derivations of Properties of Conjugate Gradient Method*

We get successively r^0, r^1, ... and p^0, p^1, ... in procedures of conjugate gradient method, then r^0, r^1, ... construct orthogonal system and p^0, p^1, ... construct A-orthogonal system. These properties guarantee that iteration terminates within the number of dimension of x.

These properties were already proved in many text-books, say [1]. We will try to derive these properties in view of our generalized gradient method.

Theorem

Series r^0, r^1, ... produced from (32)～(35) construct orthogonal system, that is

$$[r^i, r^j]=0 \quad \text{for } i\neq j, \tag{36}$$

and p^0, p^1, ... construct A-orthogonal system, that is

$$[p^i, Ap^j]=0 \quad \text{for } i\neq j. \tag{37}$$

Proof

We will use mathematical induction. From (22), (32) and (27) we have

$$[r^1, r^0]=0, \tag{38}$$

$$[p^0, Ap^1]=0. \tag{39}$$

Assume orthogonality of r^0, r^1, ... r^n and A-orthogonality of p^0, p^1, ..., p^n. Then we have

$$[r^{n+1}, r^n]=[r^{n+1}, p^n-\beta^n p^{n-1}]=0 \tag{40}$$

from (35) (24) and (26). For $i=0\sim n-1$

$$\begin{aligned}[r^{n+1}, r^i]&=[r^n-\alpha^n Ap^n, r^i] && \text{(from (34))}\\ &=-\alpha^n[Ap^n, r^i] && \text{(from the assumption)}\\ &=-\alpha^n[Ap^n, p^i-\beta^{i-1}p^{i-1}] && \text{(from (35))}\\ &=0 && \text{(from the assumption).}\end{aligned}$$

Futher from (27) we have

$$[Ap^{n+1}, p^n]=0.$$

For $i=0\sim n-1$

$$[Ap^i, p^{n+1}]=[Ap^i, r^{n+1}-\beta^n p^n] \qquad \text{(from (35))}$$
$$=[Ap^i, r^{n+1}] \qquad \text{(from the assumption)}$$
$$=[r^i-r^{i+1}, r^{n+1}]/\alpha^n \qquad \text{(from (34))}$$
$$=0 \qquad \text{(from the assumption).}$$

Thus we get orthogonality of r^0, r^1, ..., r^{n+1} and A-orthogonality of p^0, p^1, ..., p^{n+1}. q.e.d.

If the dimension of x is N, then $r^n=0$ for some $n\leqq N$ because of orthogonality of r^0, r^1, ..., r^n. So these iteration terminates at most N-th step.

7. *Discussion*

We can of couse apply our generalized gradient method to maximizing problem of general concave functions. But we have had no such actual examples at hand.

So we could say, the main pourpose of this paper is to derive conjugate gradient method from generalized gradient method the feature of which is *double Marcoff type*.

Reference

[1] RALSTON, A. AND H. S. Wilf ed. *Mathematical Methods for Digital Computers,* Wiley, 1960.