International Institute for
Applied Systems Analysis
IIASA www.iiasa.ac.at

An Accelerated Method for Minimizing a Convex Function of Two Variables

Paizerova, F.A.

IIASA Collaborative Paper
April 1985

NOT FOR QUOTATION
WITHOUT PERMISSION
OF THE AUTHOR

AN ACCELERATED METHOD FOR MINIMIZING A CONVEX FUNCTION OF TWO VARIABLES

F.A. PAIZEROVA

April 1985
CP-85-17

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
A-2361 Laxenburg, Austria

PREFACE

In this paper the author considers the problem of minimizing a convex function of two variables without computing the derivatives or (in the nondifferentiable case) the subgradients of the function, and suggests two algorithms for doing this. Such algorithms could form an integral part of new methods for minimizing a convex function of many variables based on the solution of a two-dimensional minimization problem at each step (rather than on line-searches, as in most existing algorithms.)

This is a contribution to research on nonsmooth optimization currently underway in System and Decision Sciences Program Core.

A.B. KURZHANSKI
Chairman
System and Decision
Sciences Program

AN ACCELERATED METHOD FOR MINIMIZING A CONVEX FUNCTION OF TWO VARIABLES

F.A. Paizerova

A method for minimizing a convex continuously differentiable function of two variables was proposed in [1], where it was shown that its rate of convergence is geometric with coefficient 0.9543. We shall describe two modifications of this method with improved convergence rates.

Let $Z \in E_2$, a function f be convex and continuously differentiable on E_2. Assume that we know that a minimum point of f is contained in a convex quadrilateral ABCD. The area of this quadrilateral is called the uncertainty area. Let R be the point of intersection of the diagonals of the quadrilateral. Let us choose four points M,N,Q,P on intervals AC and BD which are all at the same distance ε from R (where $\varepsilon > 0$ is fixed).

Now let us compute the function f at these points and at the point R (see Figure 1).

<u>Case 1</u>

$$f(Q) > f(R), \quad f(P) > f(R) \tag{1}$$

$$f(M) > f(R), \quad f(N) > f(R) \tag{2}$$

In this case R is (within ε-accuracy) a minimum point of f on AC and BD, and then by the properties of continuously differentiable functions the point R is a minimum point of f on ABCD (to within the given accuracy ε) and the process terminates.

<u>Case 2</u>. If inequality (1) is satisfied but inequality (2) is not, then R is a minimum point of f on BD. If $f(M) < f(R)$ then

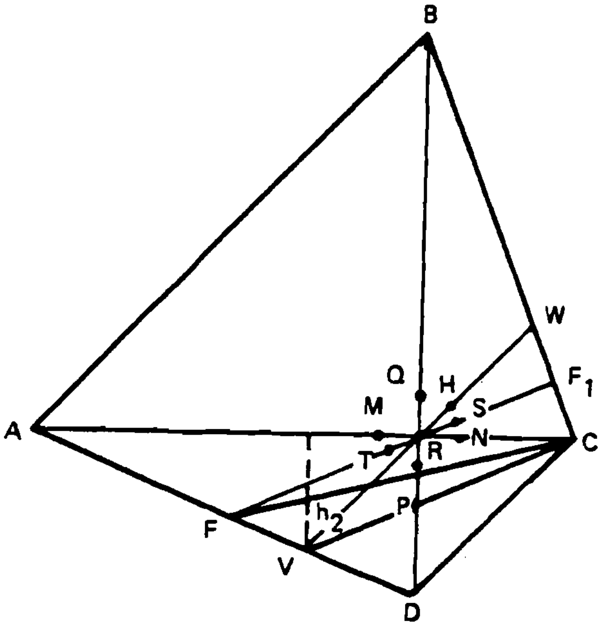

Fig. 1

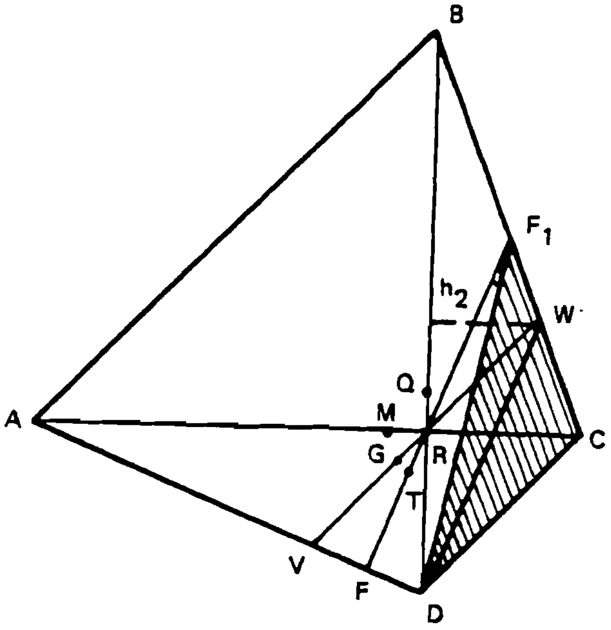

Fig. 2

$$f(Z) > f(R) \qquad \forall Z \in BDC$$

and therefore a minimum point of f lies within the triangle ABD. If $f(N) < f(R)$ then

$$f(Z) > f(R) \qquad \forall Z \in ABD$$

and a minimum point of f lies within the triangle BDC.
Case 3. If inequality (2) is satisfied but (1) is not then we argue analogously.

These three cases were discussed in [1] and are treated in the same way here. The difference between our method and that of [1] is demonstrated in the following case 4.
Case 4. Suppose that both inequalities (1) and (2) are satisfied. Then there exist two points (say, M and Q) such that

$$f(M) < f(R), \; f(Q) < f(R)$$

It follows from the convexity of f that

$$f(Z) > f(R) \qquad \forall Z \in DRC$$

Let us draw the line VW which passes through the point R and is parallel to the line DC. On the interval VW let us choose two points G and H at a distance ε from R. If $f(H) \geq f(R)$ and

$f(G) \geq f(R)$ then R is (within ε-accuracy) a minimum point of the function $f(Z)$ on the line VW (see [2]) and since $f(M) < f(R)$ then

$$f(Z) > f(R) \qquad \forall Z \in VWCD$$

This case was also discussed in [1]. The case left to be discussed is the one where either $f(H) < f(R)$ or $f(G) < f(R)$. At this point our method diverges from the method described in [1]. We will suggest two modifications of this method. For the sake of argument assume that $f(H) < f(R)$.

1. First modification. It is assumed that

$$f(H) < f(R)$$

Then (see Figure 1)

$$f(Z) > f(R) \qquad \forall Z \in VRCD$$

Moreover,

$$f(Z) > f(R) \qquad \forall Z \in VCD$$

Let us draw the line FF_1 which passes through the point R and is parallel to the line VC. On the interval FF_1 let us choose two points T and S at a distance ε from R.

If

$$f(T) \geq f(R) \quad \text{and} \quad f(S) \geq f(R)$$

then R is (within ε-accuracy) a minimum point of f on FF_1 and

$$f(Z) > f(R) \qquad \forall Z \in FF_1CD$$

If

$$f(S) < f(R) \text{ then}$$

$$f(Z) > f(R) \qquad \forall Z \in FRCD$$

and furthermore,

$$f(Z) > f(R) \qquad \forall Z \in FCD \quad .$$

As a result we get the quadrilateral ABCF which contains a minimum point of the function f. Let us compute the ratio of the areas of the quadrilaterals ABCF and ABCD.

Assume that

$$\frac{RD}{BR} = \alpha, \ \frac{AR}{RC} \geq \alpha, \ \frac{RC}{AR} = \alpha_1 \geq \alpha \quad .$$

Let h be the height of the triangle ABC. Then

$$S_{ABCD} = \frac{1}{2}\,(1+\alpha)AC\cdot h; \quad S_{ACD} = \frac{1}{2}\,\alpha\, AC\cdot h \quad ,$$

$$RC = \frac{\alpha_1}{(1+\alpha_1)}\,AC \quad .$$

Here S_{ABC} is the area of the triangle ABC. We have

$$S_{VCD} = S_{DRC} = \frac{1}{2}\,\alpha\cdot h\cdot RC = \frac{\alpha}{2}\,\frac{\alpha_1}{(1+\alpha_1)}\,AC\cdot h \quad .$$

Let us define h_2. Since

$$S_{AVC} = \frac{1}{2}\,AC\cdot h_2 \quad \text{and } S_{AVC} = S_{ACD} - S_{VCD} =$$

$$= \frac{1}{2}\,\alpha\,\cdot AC\cdot h - \frac{\alpha\cdot\alpha_1}{2(1+\alpha_1)}\,AC\cdot h = \frac{\alpha}{2(1+\alpha_1)}\,AC\cdot h$$

we have

$$h_2 = \frac{S_{AVC}}{\frac{1}{2}AC} = \frac{\alpha}{1+\alpha_1}\quad h \quad .$$

This leads to

$$S_{FVC} = S_{VRC} = \frac{1}{2}\,RC\cdot h_2 = \frac{\alpha\cdot\alpha_1}{2(1+\alpha_1)^2}\,AC\cdot h \quad ,$$

$$S_{FCD} = S_{VCD} + S_{FVC} = \frac{\alpha\cdot\alpha_1}{2(1+\alpha_1)}\,AC\cdot h + \frac{\alpha\cdot\alpha_1}{2(1+\alpha_1)^2}\,AC\cdot h =$$

$$= \frac{\alpha\cdot\alpha_1(2+\alpha_1)}{2(1+\alpha_1)^2} = AC\cdot h \quad .$$

Hence, the ratio of the area of the quadrilateral ABCF to the area of the quadrilateral ABCD is

$$1 - \frac{\alpha \cdot \alpha_1 (2+\alpha_1)}{(1+\alpha)(1+\alpha_1)^2} \quad . \tag{3}$$

Since

$\dfrac{\alpha_1 (2+\alpha_1)}{(1+\alpha_1)^2} \geq \dfrac{\alpha(2+\alpha)}{(1+\alpha)^2}$ if $\alpha_1 \geq \alpha$ this result implies

$$1 - \frac{\alpha \cdot \alpha_1 (2+\alpha_1)}{(1+\alpha)(1+\alpha_1)^2} \leq 1 - \frac{\alpha^2(2+\alpha)}{(1+\alpha)^3} \quad . \tag{4}$$

If we decrease the uncertainty area as shown in Figure 2, similar arguments lead us again to (4).

If at some step it turns out that $\frac{RD}{BR} = \alpha \leq \alpha_0$ (where α_0 will be defined later) then we draw a line passing through D and parallel to AC, and then extend AB and BD until they intersect this line (see Figure 3). Instead of the quadrilateral ABCD let us take the triangle A_1BC_1. In the case of a quadrilateral we had four lines passing through R. In the case of a triangle we take the point of intersection of its medians (the point R_1) instead of R.

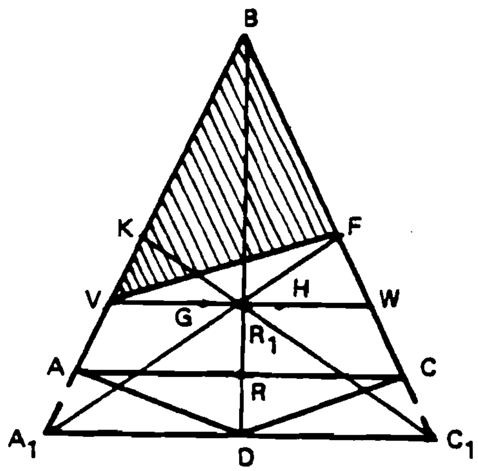

Fig. 3

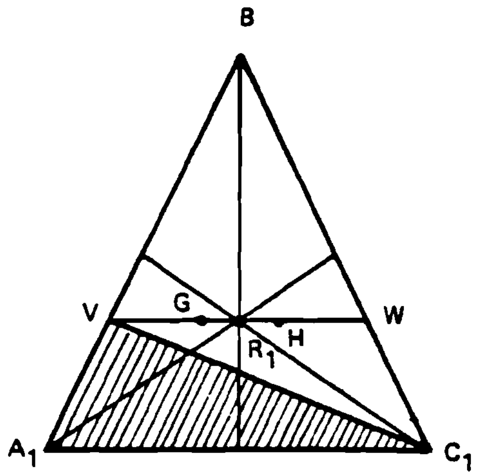

Fig. 4

If a minimum point of f is not contained in the quadrilateral $KBFR_1$ (Fig. 3) then we draw the line VW passing through R_1 and parallel to the line A_1C_1. On the interval VW let us choose two points G and H at a distance ε from R_1.

If

$$f(G) \geq f(R_1) \quad \text{and} \quad f(H) \geq f(R_1)$$

then R_1 is (within ε-accuracy) a minimum point of f on VW and

$$f(Z) > f(R_1) \qquad \forall Z \in VBW \quad .$$

Consider the case $f(H) < f(R_1)$. Then we conclude that

$$f(Z) > f(R_1) \qquad \forall Z \in VBFR_1$$

and furthermore,

$$f(Z) > f(R_1) \qquad \forall Z \in VBF \quad .$$

Thus, we have a new quadrilateral A_1VFC_1 which contains a minimum point.

Let us define the ratio of the area of the quadrilateral A_1VFC_1 and the quadrilateral ABCD. Let h be the height of the triangle ABC. We have

$$S_{ABCD} = \frac{1}{2} A_1C_1 \cdot h, \quad S_{A_1BC_1} = \frac{1}{2} (1+\alpha) A_1C_1 \cdot h \quad ,$$

$$S_{VBF} = \frac{1}{6} (1+\alpha) A_1C_1 \cdot h \quad .$$

Hence,

$$S_{A_1VFC_1} = \frac{1}{3} (1+\alpha) A_1C_1 \cdot h$$

and

$$\frac{S_{A_1VFC_1}}{S_{ABCD}} = \frac{2}{3} (1-\alpha) \quad . \tag{5}$$

Let us consider the case where the triangle $A_1R_1C_1$ (see Fig. 4) does not contain a minimum point of f. Let us draw the line VW passing through the point R_1 and parallel to the line A_1C_1, and argue as above. Let VBC_1 be a triangle which contains a minimum point of f. We get

$$S_{A_1VC_1} = \frac{1}{6}\ (1+\alpha)\ A_1C_1h$$

and the ratio of the area of the new triangle VBC_1 and the quadrilateral ABCD is $\frac{2}{3}$ $(1+\alpha)$, i.e. (5) holds again.

If $\alpha \leq \alpha_0 \sim 0.335$, then we must construct a triangle since it guarantees a greater decrease in the uncertainty area. The quantity α_0 is then a solution of the equation

$$1 - \frac{\alpha^2(2+\alpha)}{(1+\alpha)^3} = \frac{2}{3}\ (1+\alpha) \quad .$$

The convergence of this modification of the method from [1] is geometric with the rate

$$q = \frac{2}{3}\ (1+\alpha_0) \sim 0.89 \quad .$$

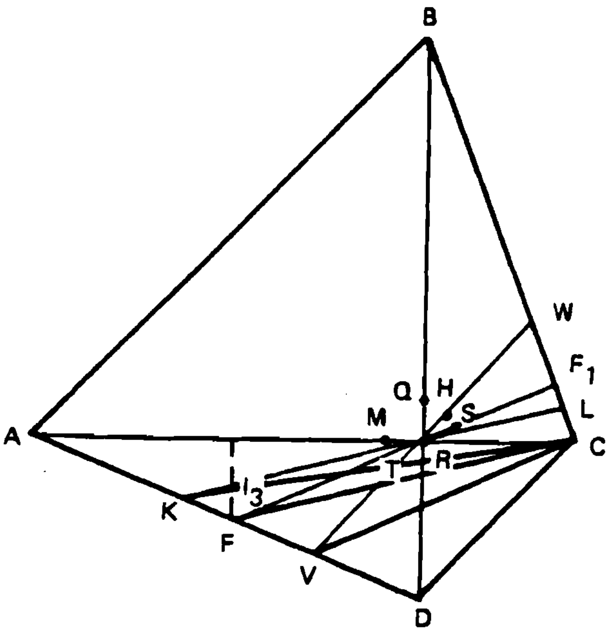

Fig. 5

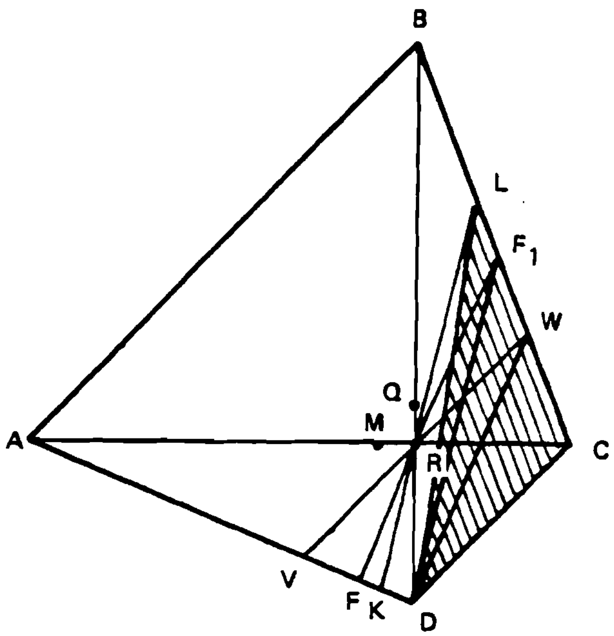

Fig. 6

2. Second modification. Let us again (see Fig. 5) assume that

$$f(M) < f(R) \quad .$$

Then

$$f(Z) > f(R) \qquad \forall Z \in ORCD \quad .$$

Furthermore,

$$f(Z) > f(R) \qquad \forall Z \in VCD \quad .$$

Let us draw the line FF_1 passing through R and parallel to the line VC. On the interval FF_1 let us choose two points T and S at a distance ε from R.

If

$$f(T) \geq f(R) \quad \text{and} \quad f(S) \geq f(R)$$

then R is (within ε-accuracy) a minimum point of f on FF_1 and

$$f(Z) > f(R) \qquad \forall Z \in FF_1CD \quad .$$

Let

$$f(S) < f(R) \quad .$$

Then

$$f(Z) > f(R) \qquad \forall Z \in FRCD$$

and furthermore

$$f(Z) > f(R) \qquad \forall Z \in FCD \quad .$$

Now let us again draw the line KL passing through R and parallel to FC and proceed as above.

As a result we get the new quadrilateral ABCK which contains a minimum point of f. Now let us compute the ratio of the areas of the new quadrilateral ABCK and the quadrilateral ABCD.

Assume that

$$\frac{RD}{BR} = \alpha, \ \frac{AR}{RC} \geq \alpha, \ \frac{RC}{AR} = \alpha_1 \geq \alpha \quad .$$

Let h be the height of the triangle ABC. It follows from the computations above that

$$S_{ABCD} = \frac{1}{2}(1+\alpha)AC\cdot h, \ S_{ACD} = \frac{1}{2}\alpha\cdot AC\cdot h \quad ,$$

$$RC = \frac{\alpha_1}{1+\alpha_1} AC, \quad S_{FCD} = \frac{\alpha\cdot\alpha_1(2+\alpha_1)}{2(1+\alpha_1)^2} AC\cdot h \quad .$$

Let us find h_3. Since $S_{AFC} = \frac{1}{2} AC\cdot h_3$ and

$$S_{AFC} = S_{ACD} - S_{FCD} = \frac{1}{2}\alpha\cdot AC\cdot h - \frac{\alpha\cdot\alpha_1(2+\alpha_1)}{2(1+\alpha_1)^2} AC\cdot h =$$

$$= \frac{1}{2}\alpha\cdot AC\cdot h\left(1 - \frac{\alpha_1(2+\alpha_1)}{(1+\alpha_1)^2}\right) = \frac{\alpha}{2(1+\alpha_1)^2} AC\cdot h$$

we have

$$h_3 = \frac{\alpha}{(1+\alpha_1)^2} AC\cdot h, \ S_{FKC} = S_{FRC} = \frac{1}{2} RC\cdot h_3 = \frac{\alpha\cdot\alpha_1}{2(1+\alpha_1)^3} AC\cdot h \ .$$

Therefore

$$S_{KCD} = S_{FCD} + S_{FKC} = \frac{\alpha\cdot\alpha_1(1+\alpha_1)}{2(1+\alpha_1)^2} AC\cdot h +$$

$$+ \frac{\alpha\cdot\alpha_1}{2(1+\alpha_1)^3} AC\cdot h = \frac{\alpha\cdot\alpha_1}{2(1+\alpha_1)^2} AC\cdot h \ (2+\alpha_1+\frac{1}{1+\alpha_1}) =$$

$$= \frac{\alpha\cdot\alpha_1(\alpha_1^2+3\alpha_1+3)}{2(1+\alpha_1)^3} AC\cdot h \quad .$$

The ratio of the areas of the new quadrilateral ABCK and the quadrilateral ABCD is

$$1 - \frac{\alpha\alpha_1(\alpha_1^2+3\alpha_1+3)}{(1+\alpha)(1+\alpha_1)^3} \quad . \tag{6}$$

Since

$$\frac{\alpha_1(\alpha_1^2+3\alpha_1+3)}{(1+\alpha_1)^3} \geq \frac{\alpha(\alpha^2+3\alpha+3)}{(1+\alpha)^3} \qquad \forall\, \alpha_1 \geq \alpha \quad ,$$

it follows from (6) that

$$1 - \frac{\alpha\alpha_1(\alpha_1^2+3\alpha_1+3)}{(1+\alpha)(1+\alpha_1)^3} \leq 1 - \frac{\alpha^2(\alpha^2+3\alpha+3)}{(1+\alpha)^4} \quad . \tag{7}$$

If we decrease the uncertainty area as shown in Fig. 6, we again obtain the same relation (7).

Let (see Fig. 7)

$$f(H) < f(R) \quad .$$

Then

$$f(Z) > f(R) \qquad \forall\, Z \in VRCD$$

and furthermore

$$f(Z) > f(R) \qquad \forall\, Z \in VCD \quad .$$

Let us draw the line FF_1 passing through the point R and parallel to the line VC. On the interval FF_1 let us choose two points T and S at a distance ε from R. If

$$f(T) \geq f(R) \quad \text{and} \quad f(S) \geq f(R)$$

then R is (within ε-accuracy) a minimum point of f on FF_1 and

$$f(Z) > f(R) \qquad \forall\, Z \in FF_1CD \quad .$$

Let

$$f(T) < f(R).$$

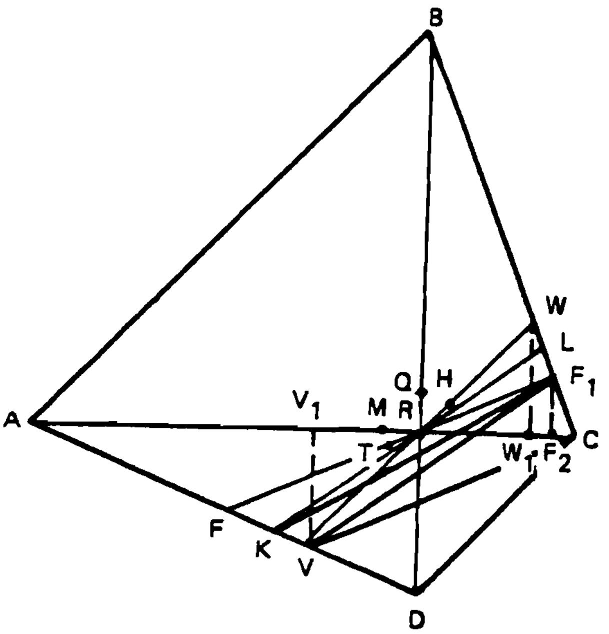

Fig. 7

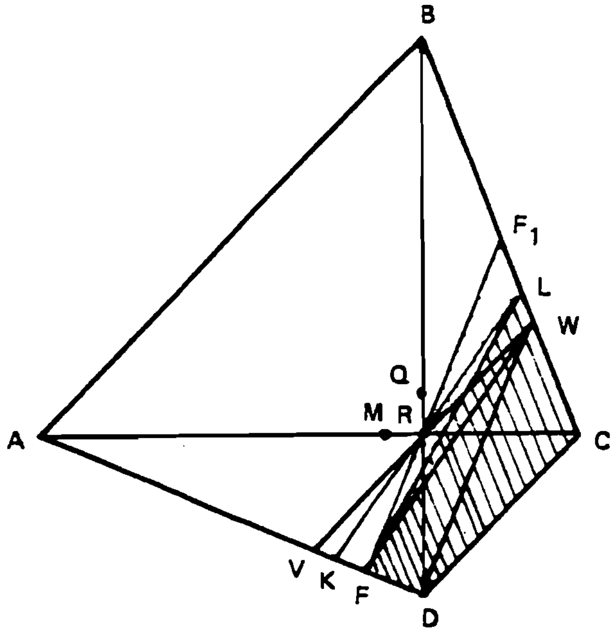

Fig. 8

Then

$$f(Z) > f(R) \qquad \forall\, Z \in VRF_1CD$$

and furthermore

$$f(Z) > f(R) \qquad \forall\, Z \in VF_1CD \quad .$$

Let us again draw the line KL passing through R and parallel to the line VF_1 and argue as above. As a result we get a new quadrilateral ABF_1K which contains a minimum point of f. Find the ratio of the areas of the quadrilaterals ABF_1K and ABCD.

Assume that

$$\frac{RD}{BR} = \alpha, \ \frac{RD}{AR} = \alpha_1 = \alpha, \ \frac{AR}{RC} \geq \alpha \quad .$$

The triangles DRC and ABR are similar since

$$\frac{RD}{BR} = \frac{RC}{AR} = \alpha, \ \angle\, DRC = \angle\, ARB \quad .$$

We have $\frac{DC}{AB} = \alpha$ and DC is parallel to AB.
The line VW is parallel to the line DC by construction. Thus, $VW \parallel AB$. The triangles ABD and VRD are also similar since the

corresponding angles are equal. Therefore

$$\frac{BD}{RD} = \frac{AB}{VR} \quad .$$

Analogously the fact that the triangles BCD and BWR are similar implies that

$$\frac{BD}{RB} = \frac{DC}{WR} \quad .$$

Therefore VR = WR and $\angle$ ARV = $\angle$ CRW. We have $VV_1 = WW_1$. The line FF_1 is parallel to the line VC by construction. Since the triangles VWC and RWF_1 are similar, we have

$$\frac{VW}{WR} = \frac{WC}{WF_1} = 2 \quad .$$

Hence,

$$WF_1 = F_1C, \; F_1F_2 = \frac{1}{2} WW_1 = \frac{1}{2} VV_1 \quad .$$

We have

$$S_{KF_1CD} = S_{VCD} + S_{VF_1C} + S_{KF_1V} = S_{VCD} + S_{VF_1C} + S_{VRF_1} =$$
$$= S_{VCD} + S_{VRC} + S_{RF_1C} \quad .$$

From the computations above it follows that

$$RC = \frac{\alpha_1}{1+\alpha_1} AC, \; VV_1 \equiv h_2 = \frac{\alpha}{1+\alpha_1} h, \; S_{VCD} = \frac{\alpha\alpha_1}{2(1+\alpha_1)} AC \cdot h \; ,$$

$$S_{VRC} = \frac{\alpha\alpha_1}{2(1+\alpha_1)^2} AC \cdot h, \; S_{ABCD} = \frac{1}{2} (1+\alpha) AC \cdot h \quad .$$

Thus,

$$S_{RF_1C} = \frac{1}{2} RC \cdot FF_1 = \frac{\alpha\alpha_1}{4(1+\alpha_1)^2} AC \cdot h \quad .$$

Then

$$S_{KF_1CD} = \frac{\alpha\alpha_1(2\alpha_1+5)}{4(1+\alpha_1)^2} = AC \cdot h \quad .$$

The ratio of the areas of the new quadrilateral ABF_1K and the quadrilateral ABCD is

$$1 - \frac{\alpha\alpha_1(2\alpha_1+5)}{2(1+\alpha_1)^2(1+\alpha)} = 1 - \frac{\alpha^2(2\alpha+5)}{2(1+\alpha)^3} \tag{8}$$

(since $\alpha_1 = \alpha$).

If we decrease the uncertainty area as shown in Fig. 8 then we again have (8). The estimate (8) is worse than (7).

In the case

$$\frac{RD}{AR} = \alpha_1 > \alpha$$

we always have an estimate better than (8). If at some step

$$\frac{RD}{BR} = \alpha \leq \alpha_0$$

then we enlarge the quadrilateral to a triangle and instead of the quadrilateral ABCD we take the triangle A_1BC_1 (Fig. 9).

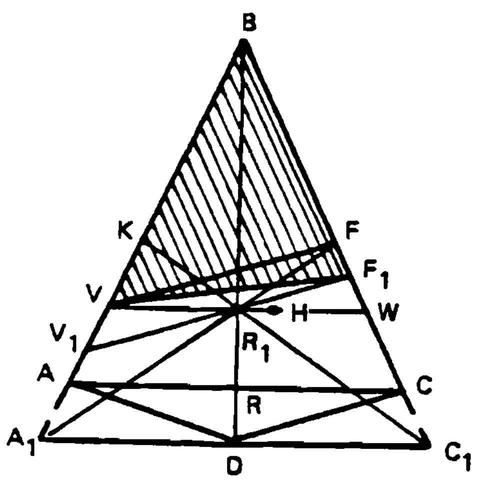

Fig. 9

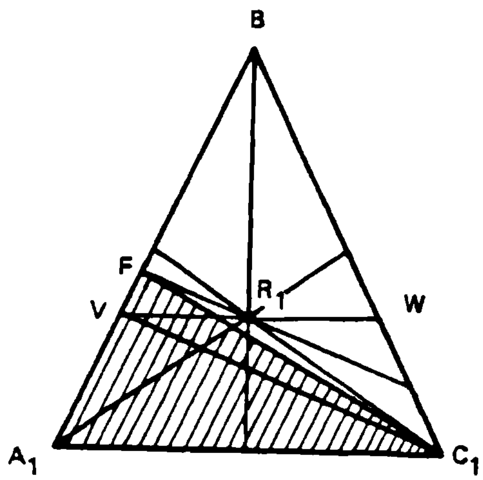

Fig. 10

Let R_1 be the point of intersection of the medians of triangle A_1BC_1. Let there be no minimum point of f in the quadrilateral $KBFR_1$. Then let us draw the line VW passing through the point R_1 and parallel to the line A_1C_1. On the interval VW choose two points G and H at a distance ε from R_1. If

$$f(G) \geq f(R_1) \quad \text{and} \quad f(H) \geq f(R_1)$$

then R_1 is (within ε-accuracy) a minimum point of f on VW and

$$f(Z) > f(R_1) \qquad \forall Z \in VBW \quad .$$

In the case $f(H) < f(R_1)$ we have

$$f(Z) > f(R_1) \qquad \forall Z \in VBFR_1$$

and moreover

$$f(Z) > f(R_1) \qquad \forall Z \in VBF \quad .$$

Let us draw the line V_1F_1 passing through the point R_1 and parallel to the line VF, and argue analogously. Let a quadrilateral $A_1VF_1C_1$ be obtained which contains a minimum point of f. Let h be the height of the triangle ABC. We have

$$S_{ABCD} = \frac{1}{2} A_1C_1 \cdot h, \; S_{A_1BC_1} = \frac{1}{2} (1+\alpha) A_1C_1 \cdot h \quad ,$$

$$S_{VBF} = \frac{1}{6} (1+\alpha) A_1C_1 \cdot h, \; S_{VFF_1} = S_{VFR_1} = \frac{1}{36} (1+\alpha) A_1C_1 \cdot h \quad ,$$

$$S_{VBF_1} = \frac{1}{37} (1+\alpha) \; A_1C_1 \cdot h \quad .$$

The ratio of the new quadrilateral $A_1VF_1C_1$ and the quadrilateral ABCD is

$$\frac{11}{18} (1+\alpha) \quad . \qquad (9)$$

If we decrease the triangle as shown in Fig. 10, then the ratio of the areas of the new triangle FBC_1 and the quadrilateral ABCD is

$$\frac{5}{9}\,(1+\alpha) \quad . \qquad (10)$$

The estimate (9) is worse than the estimate (10).

If

$$\alpha \leq \alpha_0 \approx 0.3787$$

then it is necessary to construct a triangle. The quantity α_0 is a solution of the equation

$$1 - \frac{\alpha^2(2\alpha+5)}{2(1+\alpha)^3} = \frac{11}{18}\,(1+\alpha) \quad .$$

This modification of the method displays geometric convergence with a rate $q \approx 0.8425$.

REFERENCES

1. V.F. Demyanov. "On minimizing a convex function on a plane", Zh. Vychisl. Mat. Mat. Fiz. 16(1) (1976) 247-251.

2. D.J. Wilde. Optimum Seeking Methods. Prentice-Hall Intern. Series in the Physical and Chemical Engineering Sciences, Prentice-Hall, Englewood Cliffs, N.J., 1964.