



HAL
open science

An improvement of the adaptive transformation network method by a new global minimization algorithm and its application to forecasts of air pollutions

Norio Baba, François Delebecque

► **To cite this version:**

Norio Baba, François Delebecque. An improvement of the adaptive transformation network method by a new global minimization algorithm and its application to forecasts of air pollutions. [Research Report] IRIA-RR-368, IRIA. 1979, pp.23. hal-04716626

HAL Id: hal-04716626

<https://inria.hal.science/hal-04716626v1>

Submitted on 1 Oct 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Inv: 07817 RR 368

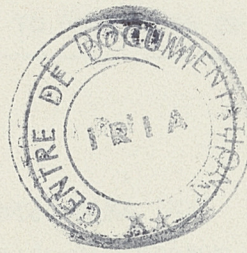
IRIA

laboria

Institut de Recherche
d'Informatique
et d'Automatique

Domaine de Voluceau
Rocquencourt
B. P. 105 78150 - Le Chesnay
France
Tél.: 954 90 20

laboratoire de recherche
en informatique
et automatique



**AN IMPROVEMENT OF THE
ADAPTIVE TRANSFORMATION
NETWORK METHOD
BY A NEW GLOBAL
MINIMIZATION ALGORITHM
AND ITS APPLICATION TO
FORECASTS OF AIR POLLUTIONS**

Norio BABA
François DELEBECQUE

Rapport de Recherche N° 368

Novembre 1979

BY A NEW GLOBAL MINIMIZATION ALGORITHM

AND ITS APPLICATION TO FORECASTS OF AIR POLLUTIONS

Norio BABA (**)

François DELEBECQUE (***)

RESUME : On propose dans ce papier une amélioration de la méthode du réseau de transformation adaptative qui peut être utilisée à la prévision à court terme de la pollution de l'air. On étudie un algorithme mixte, combinaison de méthodes d'optimisation déterministe (D.F.P. ou gradient conjugué) et d'optimisation avec recherche aléatoire.

ABSTRACT : This paper deals with an improvement of the Adaptive Transformation Network Method which is known as a useful tool for the prediction of air pollution. The improvement consists in using a new hybrid algorithm of minimization, combining the D.F.P. method (or conjugate gradient method) with the random optimization method.

(**) Tokushima University, JAPAN.

(***) IRIA/Laboria

1. - INTRODUCTION

Air pollution is one of the most serious problems for the people who live in big cities or industrial areas : This phenomenon has harmful influences upon human health and farm products. Various trials have been presented to try to give accurate forecasts of air pollution density. Among these, the Adaptive Transformation Network Method proposed by R.L. Barron [1] may be one of the most powerful tools to give good forecast of air pollution density. However, in the utilization of the Adaptive Transformation Network Method, one is faced with several difficult problems. Among these problems, one of the most important comes from the fact "It takes a lot of time to find out an appropriate parameter vector included in the Adaptive Transformation Network Method".

In this paper, we will propose a new combined algorithm of the DFP method (or conjugate gradient method) and the random optimization method. Further it is shown that the Adaptive Transformation Network Method is improved by the application of our proposed hybrid algorithm. In the first part of this paper, a brief explanation of the Adaptive Transformation Network Method is given. In the second part of this paper, we point out one of the most important problems of the Adaptive Transformation Network Method. In the third part of this paper, a hybrid algorithm of the DFP method (or conjugate gradient method) and the random optimization method is proposed in order to solve the problem of the Adaptive Transformation Network Method. Further, a convergence theorem of our hybrid algorithm is given. In order to investigate the speed of convergence, several numerical examples are also given.

2. - ADAPTIVE TRANSFORMATION NETWORK METHOD AND ITS APPLICATION TO THE FORECAST OF AIR POLLUTION DENSITY

In this section, we will give a brief introduction of the Adaptive Transformation Network Method with the examples of its application to the forecast of air pollution density.

Generally, the phenomena of air pollutions are very complicated and it is quite difficult to describe them by mathematical models because they are influenced by too many factors. And so, to give accurate forecasts of air pollutions is also very difficult.

Up to now, several trials have been presented with the object of giving good forecasts of pollutions. As shown in [3], the Adaptive Transformation Network Method proposed by R.L. Barron is one of the most powerful instruments for that object.

The Adaptive Transformation Network Method was developed in order to give an approximate inputs-output relation of the real complex system using the past data. In the following four steps, we will show how to apply the Adaptive Transformation Network Method to give good forecasts of air pollutions.

Suppose that one is faced with the problem to forecast a specific value y measuring a level of air pollution density such as SO_2 , NO_2 , dust, etc.

Step 1 : Pick out the inputs x_i ($i = 1, \dots, N$) that have significant influences upon the output y .

Step 2 : Construct a transformation network, combining such elements of 2 inputs-output relation as shown in Figure 1.

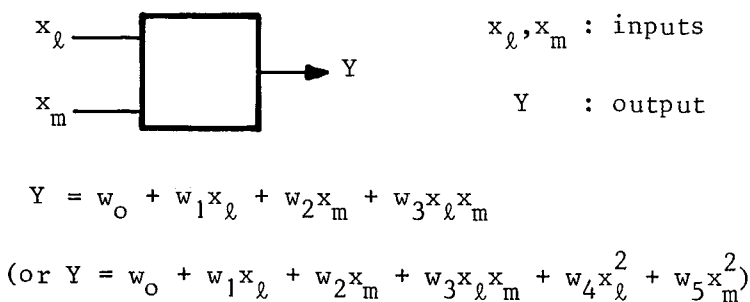


Figure 1

In Figure 2 and Figure 3, two examples of transformation networks are given.

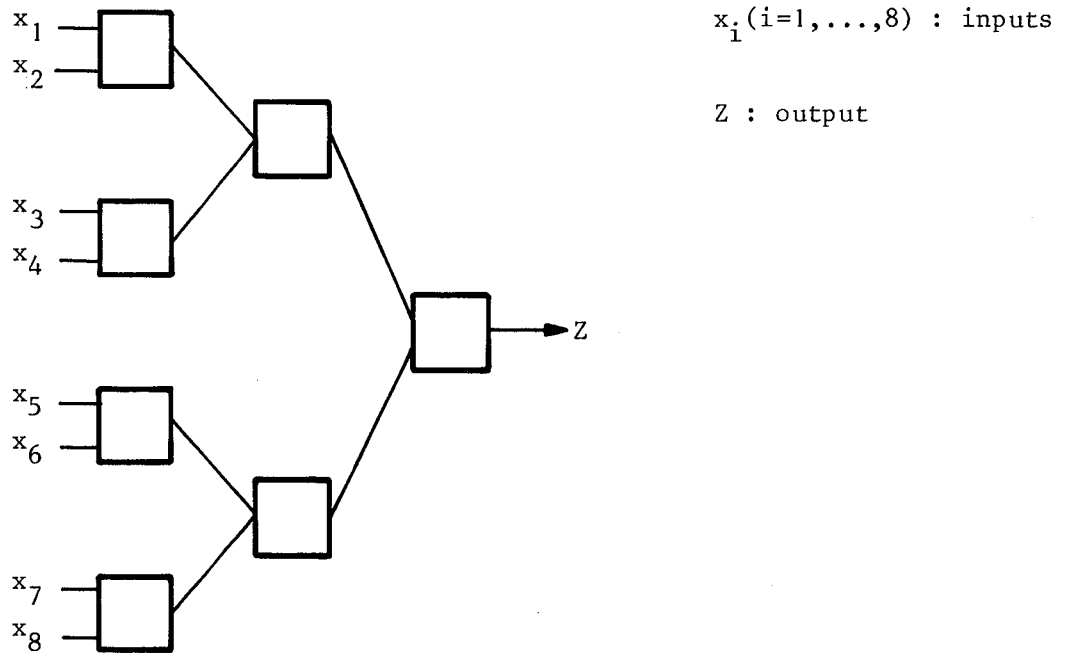


Figure 2

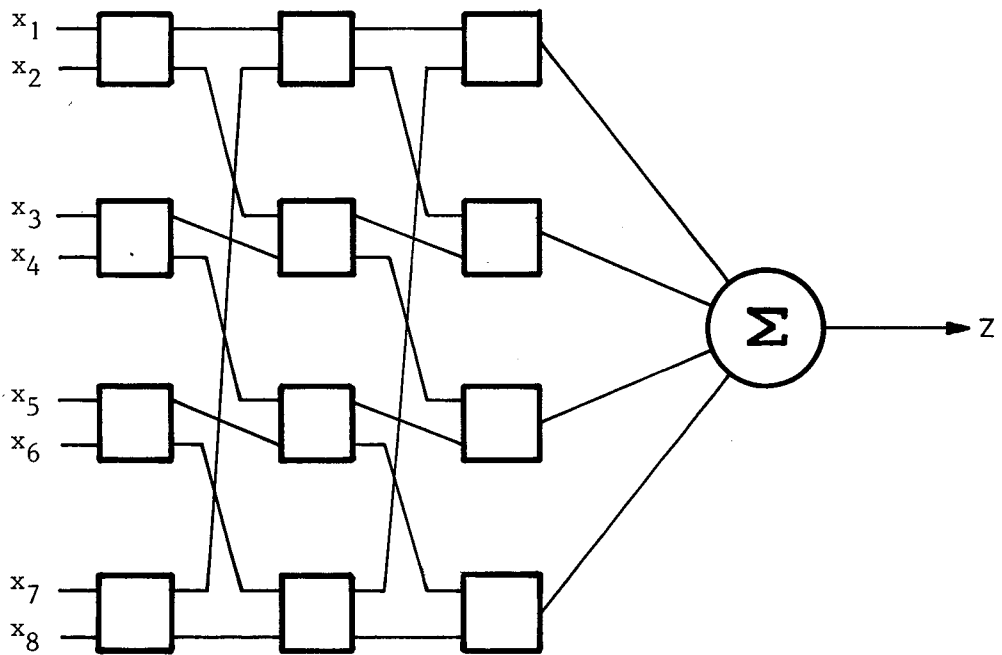


Figure 3

Step 3 : A transformation network contains a lot of unknown parameters. (For an example, the transformation network of Figure 2 contains 28 unknown parameters).

In this third step, we must find out an appropriate parameter vector \tilde{w} in order to construct a transformation network that gives a good inputs-output relation. (Further details of this step will be given in Section 3.)

R.L. Barron proposes that this procedure can be done by the random optimization method. However, as it will be mentioned in Section 3, we must do many iterations in order to find out \tilde{w} .

Step 4 : Give forecasts using the transformation network that includes the parameter vector \tilde{w} . To this purpose, we can easily implement the inputs-output relation on computer which will give a very fast response to the real inputs.

In Figure 4, forecasts of SO₂ density at noon (From November 13, 1970 to November 30, 1970 ; Tokyo, Japan) are given using the informations up to 7 o'clock [3]. In this example, we improved the parameter vector w using the data from October 22, 1970 to November 11, 1970.

In Figure 5, forecasts of SO₂ density at 3 p.m. (From June 1, 1975 to June 30, 1975 ; Tokushima, Japan) are given using the informations up to 1 p.m. 8 inputs are selected in the above two examples and the transformation network shown in Figure 2 is used. (For further details, see [3].)

3. - A PROBLEM INVOLVED IN THE ADAPTIVE TRANSFORMATION NETWORK METHOD

As shown in [3], the Adaptive Transformation Network Method is one of the most powerful tools to give good forecasts of air pollutions. However, there exist several problems to be settled. It seems that the third step (to find out an appropriate parameter vector \tilde{w}) is most difficult among the four steps in Section 2.

This step can be reduced to the nonlinear programming problem. In order to explain this briefly, let us consider the following problem :

"Give accurate forecast of SO₂ density at T o'clock using the informations of the past N days".

Let F(x,w) be the inputs-output relation of the Adaptive Transformation Network.

(It should be noted that it is rather difficult to write the explicit form of the F functions. However, the computation can be done easily in a recursive way by computing the outputs of each elements of the network and then entering these outputs as inputs for other elements.)

Let \bar{Z}_j, \bar{x}_j and w be SO₂ density at T o'clock of the j-th day, input vector of the j-th day, and parameter vector included in the Adaptive Transformation Network respectively.

Now, let us consider the following objective function :

$$G(w) \hat{=} \sum_{j=1}^N (\bar{Z}_j - F(\bar{x}_j, w))^2$$

In the third step, we must find out an appropriate parameter \tilde{w} that gives small value of G(w). (This means that the third step can be reduced to the nonlinear minimization problem.) However, G(w) is not necessarily a unimodal performance function : G(w) admits several local minima, and therefore, an ordinary descent algorithm will generally lead to a local minimum. R.L. Barron uses the random optimization method in order to find out the global minimum. Although it is shown that the global minimum is found out by the Matyas' random optimization method [5] without regard to the selection of the initial point [6], it takes a lot of time to find out the global minimum as shown by the numerical examples of the next section. In the next section, we propose a new combined algorithm that ensures convergence to a global minimum and gives a faster rate of convergence.

SO₂ density (pphm)

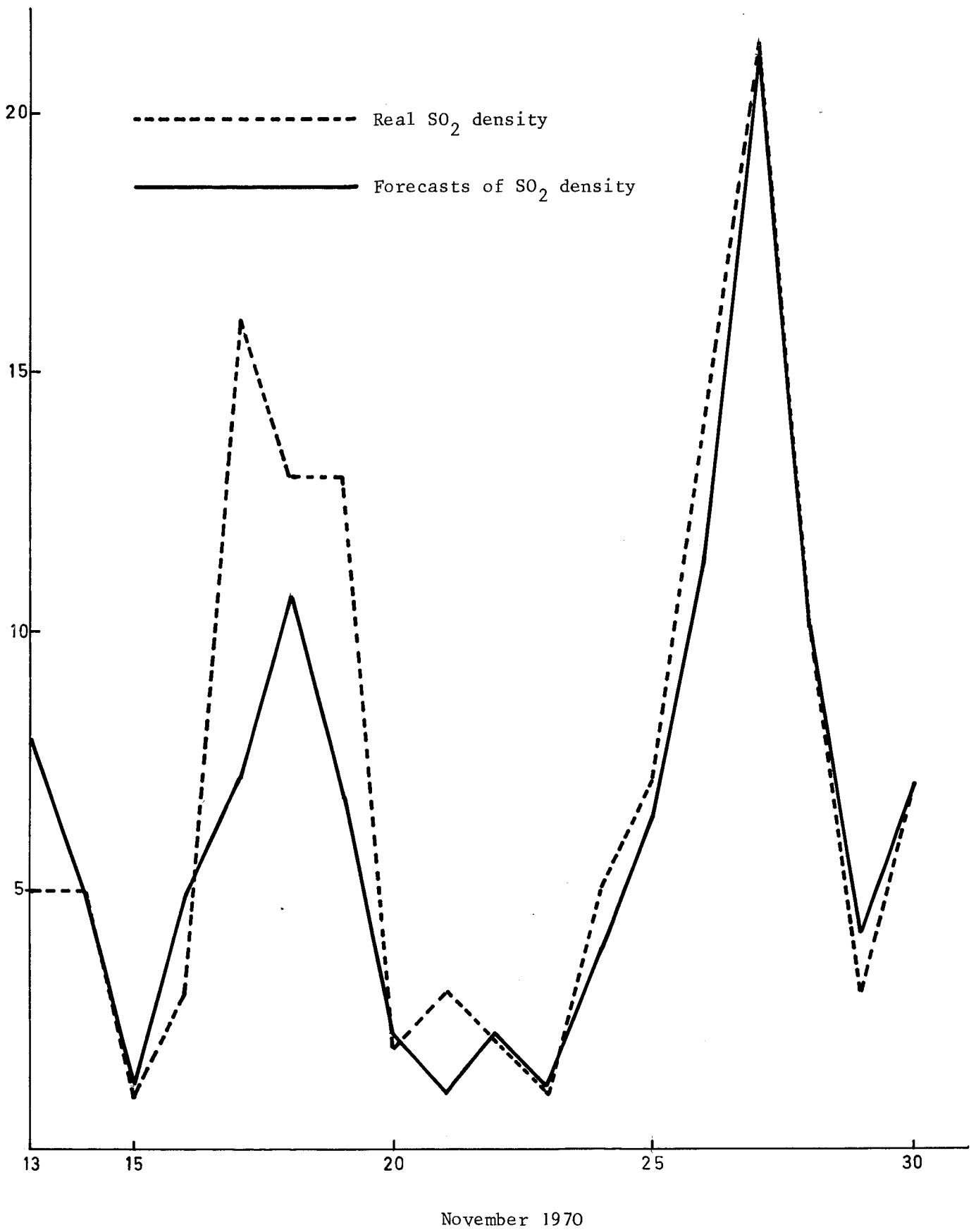
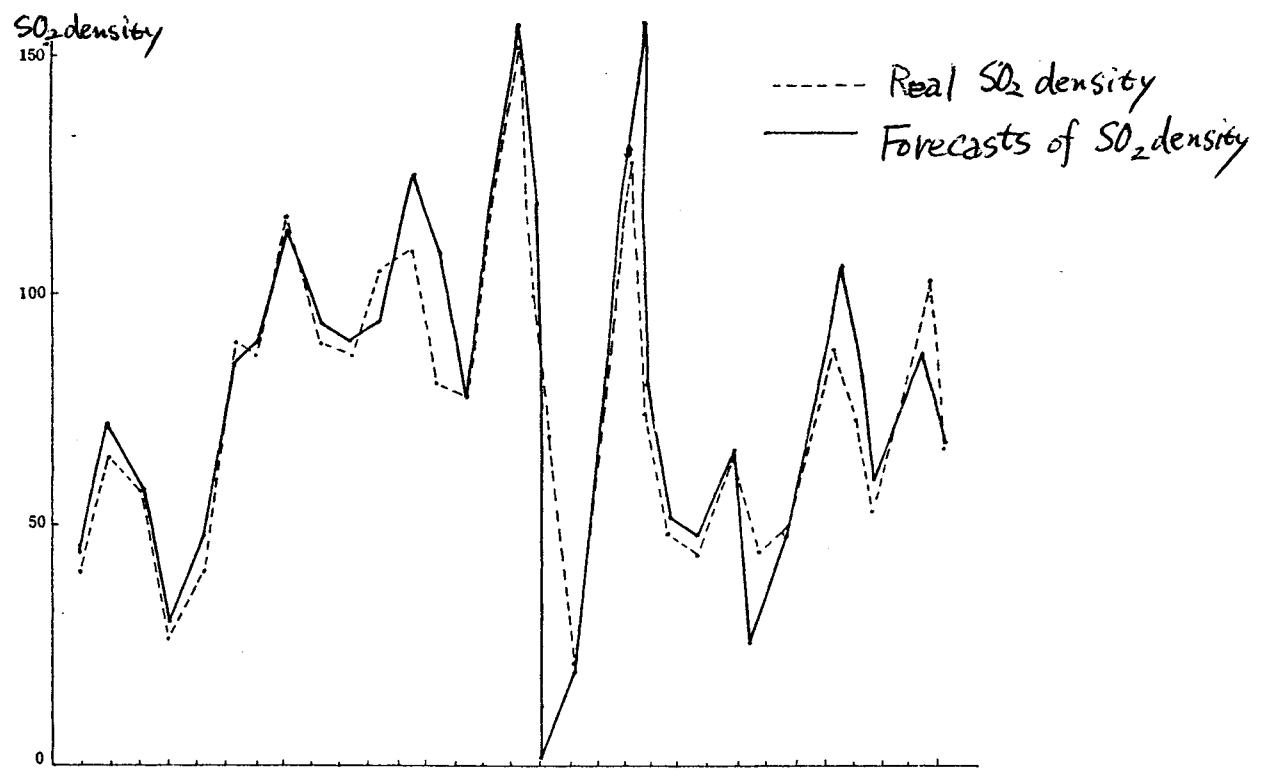


Figure 4 : Forecasts of SO₂ density at noon



June 1975

Figure 5 : Forecasts of SO₂ density at 3 p.m.

4. - A HYBRID ALGORITHM TO FIND THE GLOBAL MINIMUM AND ITS UTILIZATION TO IMPROVE THE ADAPTIVE TRANSFORMATION NETWORK METHOD

In this section, we describe a new combined descent algorithm of the random optimization method and the DFP method (or conjugate gradient method) in order to find out the global minimum of the objective function $f(x)$, $x \in S$, S : the subset of E^n , in a small number of steps. First of all, let us briefly refer to the random optimization method and the DFP method. (As to the conjugate gradient method, see [9] and [11].)

4.a.1. - Random Optimization Method

The random optimization method was originally proposed by Rastrigin [4]. After that, various modified algorithms of his original random optimization method have been contrived.

The following algorithm is known as Matyas' random optimization method [5][6].

Let $\xi^{(i)}$ be n -dimensional random vectors with zero mean ($i=1, \dots$) and have the following property.

There exists some positive number θ such that :

$$\theta = \inf_{x, y \in S} q(x-y) > 0$$

(Here, $q(\cdot)$ is the probability density function of $\xi^{(i)}$ ($i=1, \dots$)).

Step 1 : Select an initial point $x^{(1)} \in S$ and compute $f(x^{(1)})$

Step 2 : The next point is determined as follows.

If $x^{(1)} + \xi^{(1)} \in S$, compute $f(x^{(1)} + \xi^{(1)})$.

If $f(x^{(1)} + \xi^{(1)}) < f(x^{(1)})$, $x^{(2)} \triangleq x^{(1)} + \xi^{(1)}$.

If $f(x^{(1)} + \xi^{(1)}) \geq f(x^{(1)})$, $x^{(2)} \triangleq x^{(1)}$.

If $x^{(1)} + \xi^{(1)} \notin S$, let $x^{(2)} \triangleq x^{(1)}$.

Step 3 : In the general k-th step, $x^{(k+1)}$ is determined as follows. ($k= 1, \dots$)

If $x^{(k)} + \xi^{(k)} \in S$, compute $f(x^{(k)} + \xi^{(k)})$.

If $f(x^{(k)} + \xi^{(k)}) < f(x^{(k)})$, $x^{(k+1)} \triangleq x^{(k)} + \xi^{(k)}$.

If $f(x^{(k)} + \xi^{(k)}) \geq f(x^{(k)})$, $x^{(k+1)} \triangleq x^{(k)}$.

If $x^{(k)} + \xi^{(k)} \notin S$, let $x^{(k+1)} \triangleq x^{(k)}$.

This algorithm converges to the global minimum without regard to the location of the initial point [6].

4.a.2. - DFP method

The DFP (Davidon-Fletcher-Powell) Method is defined as follows.

Step 1 : Select an initial point $x^{(1)} \in S$ and a symmetric positive definite matrix S_1 .

Let $k = 1$.

Step 2 : Set $d_k = -S_k g_k$ ($g_k = \nabla f(x^{(k)})'$).

Step 3 : Minimize $f(x^{(k)} + \alpha d_k)$ to get $x^{(k+1)}$.
 $\alpha \geq 0$

$x^{(k+1)} \triangleq x^{(k)} + \alpha_k d_k$, $P_k \triangleq \alpha_k d_k$,

$g_{k+1} \triangleq \nabla f(x^{(k+1)})'$, and $q_k \triangleq g_{k+1} - g_k$.

Step 4 :

$$S_{k+1} \triangleq S_k + \frac{P_k P_k'}{P_k' q_k} - \frac{S_k q_k q_k' S_k}{g_k' S_k q_k}$$

Let $k = k+1$ and return to step 2.

This algorithm converges in a finite number of steps for a positive definite quadratic problem. For non-quadratic convex objective functions this algorithm has a nice property of n-step quadratic convergence [10].

4.b. - Combined Algorithm

We will propose a new combined algorithm that makes use of both merits of the random optimization method and the DFP (or conjugate gradient) method.

An outline of this hybrid algorithm is shown in Figure 6.

Let us explain it briefly. First, minimization procedure is carried out using the DFP method. When the decrease of the value of the objective function becomes smaller than a specified value ϵ_1 , we change the overall descent algorithm from the DFP method to the random optimization method in order to prevent it from stopping at a local minimum point. If the decrease of the value of the objective function becomes larger than a specified value G , we change the overall descent algorithm from the random optimization method to the DFP method. (The object of this switching is to speed up our minimization process.) The same procedures as above are used repeatedly.

When the total number of steps exceeds a specified number M , the overall descent algorithm is stopped.

The real problem can often be stated as "Find the smallest possible value of a performance function within a given number of steps M ." Probably, our proposed hybrid algorithm is suitable to solve such problems where the performance function admits several local minima.

Remark : In order to find out an appropriate parameter value of the Adaptive Transformation Network Method, we must generally deal with multi-dimensional problems. In such a case, the following practical method can be used to execute our overall descent algorithm, when the computer capacity is not sufficient.

Let $w^{(k)}$ be the parameter vector at the k -th step of the transformation network.

$$w^{(k)} \triangleq (w_1^{(k)}, \dots, w_m^{(k)})'$$

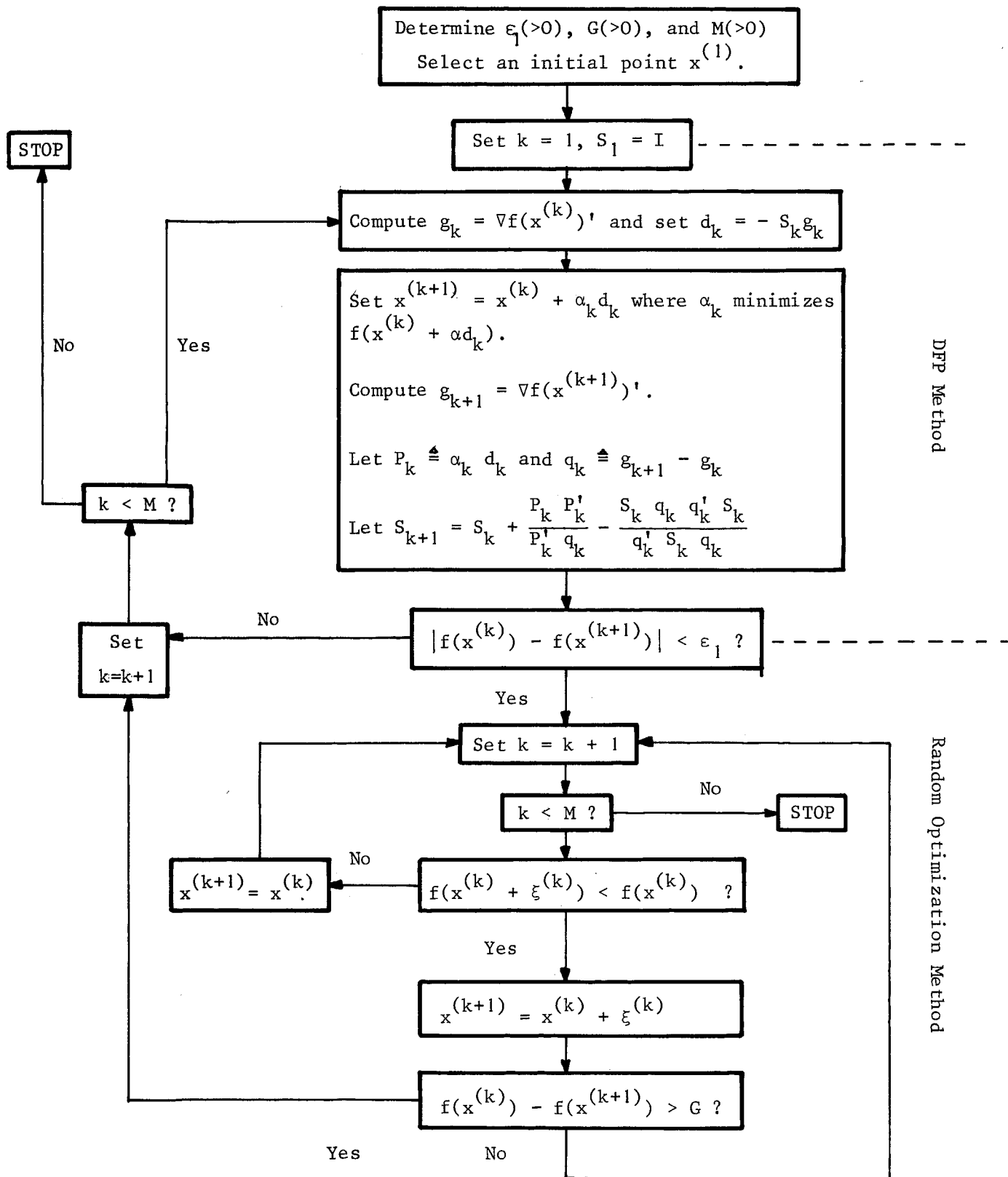


Figure 6 : Combined Algorithm of the DFP method and the Random Optimization method

Further, let us assume that $w^{(k)}$ was found by the random optimization method and the $(k+1)$ -th step must be carried out to get $w^{(k+1)}$ using the DFP method. (See Figure 6). In this case, we only execute the $(k+1)$ -th step of the minimization procedure concerning some (L) of the elements of the parameter vector $w^{(k)}$ that correspond to the larger changes in the k -th step. ($L < m$)

(Suppose that :

$$\begin{aligned} |w_{\alpha_1}^{(k)} - w_{\alpha_1}^{(k-1)}| &\geq |w_{\alpha_2}^{(k)} - w_{\alpha_2}^{(k-1)}| \geq \dots \geq |w_{\alpha_L}^{(k)} - w_{\alpha_L}^{(k-1)}| \\ &\geq \dots \geq |w_{\alpha_m}^{(k)} - w_{\alpha_m}^{(k-1)}| \end{aligned}$$

$$(\alpha_i \neq \alpha_j \text{ (} i \neq j \text{)} \mid \leq \alpha_t \leq m \text{ (} t=1, \dots, m \text{))}.$$

In this case, we fix the parameters $w_{\alpha_{L+1}}^{(k)}$, ..., and $w_{\alpha_m}^{(k)}$ and execute the $(k+1)$ -th step of the minimization procedure concerning the elements $w_{\alpha_1}^{(k)}$, ..., and $w_{\alpha_L}^{(k)}$. The number L must be decided from the point of view of computer capacity, etc.)

4.c. - Experimental Results

In this section, we consider the following three examples and compare our proposed algorithm with the random optimization method and the DFP method (or conjugate gradient method).

Example (i) : Minimize the following function :

$$f(x) = x_1^2 + 2x_1x_2 + 2x_2^2 + 6x_1 + 8x_2 + 1 \quad \text{for } x_1 \geq -4$$

$$f(x) = 2x_1^2 + 2x_1x_2 + x_2^2 + 16x_1 + 4x_2 + 5 \quad \text{for } x_1 < -4$$

The above objective function is a multimodal function that has the local minimum -9 at the point $(-2, -1)$ and the global minimum -35 at the point $(-6, 4)$.

Let $\epsilon_1 \hat{=} 0.003$, $G \hat{=} 2$, and $M \hat{=} 50$.

Further, we take $(23, 25)$ as the initial point of this problem. The computational results are shown in table 1.

Table 1 : Value of the performance function
(Initial point (23, 25))

Step N°	HA	ROM	DFP
1	3268.00	3268.00	3268.00
2	7.08	3213.86	7.08
3	- 9.00	3191.30	- 9.00
4	- 9.00	3129.22	
5	*- 9.00	3129.22	
10	-11.79	2976.81	
11	*-28.93	2848.55	
12	-35.00	2848.55	
13	-35.00	2848.55	
20	-35.00	2429.63	
50	-35.00	819.29	
100		-30.74	
150		-34.97	

HA : Hybrid Algorithm of the DFP method and the Random Optimization Method
(* change),

DFP: Davidon Fletcher Powell Method,

ROM: Random Optimization Method.

Example (ii) : Minimize the following function

$$f(x) = (x_1 - 2)^2 + 3(x_2 - 2)^2 - 18 \quad \text{for } x_1 \geq 0, x_2 \geq -x_1$$

$$f(x) = 2x_1^2 + 2x_1x_2 + x_2^2 + 16x_1 + 4x_2 + 10 \quad \text{for } x_1 < 0, x_2 \geq 0.2x_1$$

$$f(x) = 4x_1^2 - 2x_1x_2 + 2x_2^2 + 4x_1 + 2x_2 - 10 \quad \text{for another region.}$$

The above function has the local minima -18 and -12.28 at the points (2, 2) and (-5/7, -6/7) respectively.

It has the global minimum -30 at the point (-6, 4). Let (-10, -30) be the initial point and $\epsilon_1 \hat{=} 0.003$, $G \hat{=} 2$, $M \hat{=} 50$.

The computational results of the above problem are shown in Table 2.

Example (iii) : Minimize the following function

$$f(x) = 5x_1^2 - 6x_1x_2 + 2x_2^2 - 12x_1 + 12x_2 + 36 \\ + 1000 \sin(\pi(x_1 + 1)/10) + 5 \cos(\pi(x_2 - 8)/20).$$

The above function has the global minimum -1005 at the point (-6, -12). We choose (-50, 50) as the initial point and let $\epsilon_1 \hat{=} 0.003$, $G \hat{=} 2$, and $M \hat{=} 50$.

The computational results of the above problem are shown in Table 3.

4.d. - Theoretical Considerations of the Proposed Hybrid Algorithm

In the preceding section, several numerical examples that indicate the effectiveness of our proposed hybrid algorithm were given. However, our proposed hybrid algorithm would become meaningless if we could not give any theoretical results that ensure convergence to the global minimum.

Table 2 : Value of the performance function

(Initial point (-10, -30))

Step N°	HA	ROM	DFP
1	1490.00	1490.00	1490.00
2	68.33	1490.00	68.33
3	- 12.29	1472.41	- 12.29
4	- 12.29	1472.41	
5	*- 12.29	1472.41	
10	- 12.29	1284.09	
14	- 12.29	1106.82	
15	- 17.40	1087.03	
16	*- 18.00	1087.03	
29	- 18.00	980.91	
30	- 20.51	905.54	
48	- 20.51	492.20	
49	- 26.78	492.20	
50	*- 29.94	492.20	
51	(- 30.00)	(492.20)	
60		153.41	
100		0.97	
200		- 29.99	

Table 3 : Value of the performance function
(Initial point (-50, 50))

Step N°	HAD	HAC	DFP	ROM
1	33431.73	33431.73	33431.73	33431.73
2	- 789.28	- 789.28	- 789.28	33336.01
3	- 792.33	- 792.79	- 792.33	33336.01
4	- 801.85	- 795.29	- 801.85	33336.01
5	- 801.85	- 801.83		31666.62
6	*- 801.85	- 801.83 *- 801.83		
21	- 801.85			
22	- 900.84			
23	*- 944.05			
24	-1004.996			
25	-1005.00			
26	*-1005.00			
35		- 801.83		23168.20
36		- 932.91		23168.20
37		*- 953.58		23168.20
38		-1005.00		23168.20
39		-1005.00		
40		*-1005.00		
50	-1005.00	-1005.00		21089.64
100				8869.63
300				-1003.53

HAD : Hybrid Algorithm of the DFP Method and the Random Optimization Method

HAC : Hybrid Algorithm of the Conjugate Gradient Method and the Random Optimization Method.

In the following, we will give some convergence result concerning our proposed hybrid algorithm.

Theorem

Let $f(x)$ be a C^1 function, defined on a subset S of E^n .

Let $\hat{x} \in S$ be the point such that

$$f(\hat{x}) = \min_{x \in S} f(x)$$

Further, let R_ϵ be a region such that :

$$R_\epsilon = \{x \mid |f(x) - f(\hat{x})| < \epsilon, x \in S\}.$$

Assume the following.

For any positive number δ , the Euclidean measure of $U_\delta(\hat{x}) \cap S$ is positive.

$$(U_\delta(\hat{x}) = \{x \mid \|x - \hat{x}\| < \delta\})$$

Further, assume that we can calculate $x^{(i)}$ ($i=1, \dots$) infinitely. (Let M be infinity.)

Then, for any positive number ϵ , the sequence $\{x^{(k)}\}$ obtained by the above algorithm converges with probability 1 to the region R_ϵ .

$$\lim_{k \rightarrow \infty} P\{\omega \mid x^{(k)} \in R_\epsilon\} = 1.$$

[Proof]

Let us consider the process with which the point $\{x^{(k)}\}$ is sought. In particular, let us consider the case when the DFP method is used.

We can consider the following three cases (a), (b), and (c).

(a) In the first step, we use the DFP method in order to get $x^{(2)}$ from $x^{(1)}$.

(b) The preceding research is done by the DFP method and the absolute value of the decrease of the objective function is larger than ε_1 .

(c) The preceding research is done by the random optimization method and the absolute value of the decrease of the objective function is larger than G .

Let N_a , N_b , and N_c be the numbers of times of occurrences of the case (a), the case (b), and the case (c) respectively.

Further, let :

$$(1) \quad t_1 \varepsilon_1 \triangleq f(x^{(1)}) - f(\hat{x})$$

$$(2) \quad t_2 G_2 \triangleq f(x^{(1)}) - f(\hat{x})$$

$$(3) \quad h \triangleq [t_1]$$

$$(4) \quad S \triangleq [t_2]$$

,where $[t_i]$ means the largest integer which is smaller than t_i ($i=1,2$).

$$(5) \quad \text{Clearly, } N_a = 1$$

From the definition of (1) ~ (4) and the fact that $f(x^{(i)})$ ($i=1, \dots$) is a monotone decreasing function of i , the following relations hold.

$$(6) \quad N_b \leq h + 1$$

$$(7) \quad N_c \leq S + 1$$

The random optimization method is used at the time when the DFP method is not used. Therefore, it becomes clear that the random optimization method is used more than T times to get $x^{(k)}$, where :

$$(8) \quad \begin{aligned} T &= (k-1) - (1 + (h+1) + (S+1)) \\ &= k - (h+S+4). \end{aligned}$$

Since $(h+S+4)$ is bounded, T tends to infinity if k tends to infinity.

Now, we will prove the above convergence theorem using the convergence property of the random optimization method.

Since $f(x)$ is a continuous function, there exists $\hat{\delta} > 0$ such that the following relation holds.

$$(9) \quad \text{If } ||x-\hat{x}|| < \hat{\delta}, \text{ then } |f(x)-f(\hat{x})| < \varepsilon$$

Let us consider the following region A .

$$(10) \quad \begin{aligned} A &\triangleq \{x \mid ||x-\hat{x}|| < \hat{\delta}, x \in S\} \\ & (= U_{\hat{\delta}}(\hat{x}) \cap S) \end{aligned}$$

From (9) and the definition of R_{ε} ,

$$(11) \quad R_{\varepsilon} \supset A$$

Let :

$$(12) \quad \theta \triangleq \inf_{x,y \in S} q(x-y)$$

Now, assume that the $(i+1)$ -th point $x^{(i+1)}$ will be found out by the random optimization method. And, let $P_A\{x^{(i)}\}$ be the probability with which the $(i+1)$ -th point $x^{(i+1)}$ enters into the region A . ($i=1,2,\dots$)

Then, this probability will be estimated as follows

$$(13) \quad P_A\{x^{(i)}\} \geq \theta M(A) \quad \text{for all } x^{(i)} \in S$$

$$(14) \quad , \text{ where } M(A) \text{ is the measure of } A \text{ in } E^n.$$

$$(15) \quad \text{Let } \gamma \text{ be } \theta M(A).$$

From the assumption of the theorem, $M(A)$ is positive. Then,

$$(16) \quad 0 < \gamma \leq 1$$

Let $P_{R_\epsilon}\{x^{(i)}\}$ be the probability with which the $(i+1)$ -th point $x^{(i+1)}$ enters into the region R_ϵ . ($i=1,2,\dots$)

$$(17) \quad \text{Since } R_\epsilon \supset A, P_{R_\epsilon}\{x^{(i)}\} \geq P_A\{x^{(i)}\}.$$

From (13), (15), and (16),

$$(18) \quad P_{R_\epsilon}\{x^{(i)}\} \geq \gamma > 0$$

If the point $x^{(j)}$ enters into the region R_ϵ , $\{x^{(t)}\}$ ($t=j+1, j+2, \dots$) do not leave this region. ($f(x^{(i)})$ is a monotone decreasing function of i .)

Therefore, the region R_ϵ is an absorbing region.

Consequently ,

$$(19) \quad P\{x^{(k)} \in R_\epsilon\} \geq 1 - (1-\gamma)^T$$

$$(T = k - (h+S+4) \quad (8))$$

Hence,

$$\lim_{k \rightarrow \infty} P\{x^{(k)} \in R_\epsilon\}$$

$$\geq \lim_{k \rightarrow \infty} \{1 - (1-\gamma)^T\}$$

$$= 1$$

Aknowledgements

This work was achieved within the collaboration between IRIA-Laboria and the Department of Applied Mathematics and Physics of Kyoto University. We would like to take the opportunity to thank Prof. Lions (IRIA), Prof. Bensoussan (IRIA), and Prof. Sawaragi (Kyoto University) for permitting us to work together.

REFERENCES

- [1] R.L. Barron, "Adaptive Transformation Networks for Modelling, Prediction, and Control", 1971 IEEE Systems, Man, and Cybernetics Group Annual Symposium Record, October 21, pp. 254-263, 1971.
- [2] A.G. Ivakhnenko, "Polynomial Theory of Complex System", IEEE Trans. Systems, Man, and Cybernetics, Vol. SMC-1, pp. 368-378, 1971.
- [3] N. Baba, T. Soeda, and Y. Sawaragi, "Forecasts of Air Pollutions Using Adaptive Transformation Method", Proceedings of IFAC Symposium, 1976, USSR, pp. 67-79.
- [4] L.A. Rastrigin, "The convergence of the random search method in the extremal control of a many parameter system", Automation and Remote Control, Vol. 24, pp. 1337-1342, 1963.
- [5] J. Matyas, "Random optimization", Automation and Remote Control, Vol. 28, pp. 246-253, 1965.
- [6] N. Baba, T. Shoman, and Y. Sawaragi, "A Modified Convergence Theorem for a Random Optimization Method", Information Sciences, Vol. 13, pp: 159-166, 1977.
- [7] W.C. Davidon, "Variable Metric Method for Minimization", A.E.C. Res. and Develop. Rept. ANL-5990. Argonne Nat'l Lab., Argonne, Illinois, 1959.
- [8] R. Fletcher and M.J.D. Powell, "A rapidly Convergent Descent Method For Minimization", The Computer Journal, Vol. 6, pp. 163-168, 1963.
- [9] R. Fletcher and C.M. Reeves, "Function Minimization by Conjugate Gradients", The Computer Journal, Vol. 7, pp. 149-154, 1964.
- [10] W. Burmeister, "Die Konvergenzordnung des Fletcher-Powell-Algorithmus", ZAMM 53, pp. 693-699, 1973.
- [11] D.G. Luenberger, "Introduction to Linear and Nonlinear Programming", Mass. Addison-Wesley Publishing Company, 1973.