APPLICATIONES MATHEMATICAE
27,2 (2000), pp. 127–141

B. RHANIZAR (Rabat and Marrakech)

ON HENRICI'S TRANSFORMATION IN OPTIMIZATION

Abstract. Henrici's transformation is a generalization of Aitken's Δ^2-process to the vector case. It has been used for accelerating vector sequences. We use a modified version of Henrici's transformation for solving some unconstrained nonlinear optimization problems. A convergence acceleration result is established and numerical examples are given.

1. Introduction. It is well known that extrapolation methods can be used to accelerate the convergence of vector sequences [2, 9, 11]. We consider Henrici's transformation introduced in [7]. This is a natural generalization of Aitken's Δ^2-process [3], and it has been used in the vector case. In this paper, we shall use a modified Henrici transformation for solving a nonlinear optimization problem. Results on convergence acceleration will be established.

Let us introduce the setting used throughout this paper. The vector $x = (x_1, \ldots, x_p) \in \mathbb{R}^p$ denotes a p-dimensional vector, and we shall use the Euclidean norm $\|x\| = (\sum_{1\le i\le p} x_i^2)^{1/2}$. We denote by $\langle\ ,\ \rangle$ the corresponding inner product. We shall also use the matrix norm $\|A\| = \sup_{\|x\|\neq 0} \|Ax\|/\|x\|$ for any matrix A.

Consider the following optimization problem:

$$\text{find } x^* \in \mathbb{R}^p \text{ such that } f(x^*) = \min_{x\in\mathbb{R}^p} f(x), \tag{1.1}$$

where f is a convex function from $\mathbb{R}^p$ to $\mathbb{R}$. The gradient method with an optimal step [1, 4, 12] is defined by the sequence

$$x_{n+1} = x_n - \lambda_n \nabla f(x_n), \qquad n = 0, 1, \ldots, \tag{1.2}$$

$$f(x_n - \lambda_n \nabla f(x_n)) = \min_{\lambda\in\mathbb{R}} f(x_n - \lambda \nabla f(x_n)), \tag{1.3}$$

where ∇f is the gradient of f. Under certain assumptions, the sequence

1991 *Mathematics Subject Classification*: Primary 49M45.
Key words and phrases: Henrici's transformation, nonlinear optimization.

(x_n) defined by (1.2) converges to the solution of problem (1.1) (see [4]). Unfortunately, the convergence is so slow that the method is of no practical use. Thus, in such cases, it is fundamental to accelerate the convergence of (x_n).

We recall the Henrici transformation and propose its modified version for accelerating the convergence of (x_n). The *Henrici transformation* [11] with respect to the sequence (x_n) amounts to considering the sequence (h_n) given by

$$h_n = x_n - \Delta X_n(\Delta^2 X_n)^{-1}\Delta x_n, \qquad n = 0, 1, \ldots, \tag{1.4}$$

where ΔX_n denotes the $p\times p$ matrix whose columns are $\Delta x_n, \ldots, \Delta x_{n+p-1}$, with $\Delta x_k = x_{k+1} - x_k$ for $k = n, \ldots, n+p-1$, and where $\Delta^2 X_n$ is the matrix whose columns are $\Delta^2 x_n, \ldots, \Delta^2 x_{n+p-1}$, with $\Delta^2 x_k = \Delta x_{k+1} - \Delta x_k$ for $k = n, \ldots, n+p-1$.

The modified version of (1.4) that we propose is as follows:

$$h'_n = x_n - \Delta X_n(\Delta F'(x_n))^{-1}\nabla f(x_n), \qquad n = 0, 1, \ldots, \tag{1.5}$$

where $\Delta F'(x_n)$ is the $p \times p$ matrix whose columns are

$$\nabla f(x_{n+1}) - \nabla f(x_n), \ldots, \nabla f(x_{n+p}) - \nabla f(x_{n+p-1}).$$

For this choice we give a convergence acceleration result (Theorem 2.1). The remaining part of the paper is organized as follows. Section 2 is devoted to results on convergence acceleration by using the transformation (h'_n) given by (1.5). Illustrative numerical examples are considered in Section 3.

2. Convergence acceleration results. Consider the sequence $(x_n)_n$ defined by (1.2). In this section we will first show that (h'_n) is well defined and converges faster than (x_n), secondly we will show that (h'_n) accelerates (x_{n+p}). A comparison between h'_{n+1} and h'_n will also be given. We denote by

$$G(u_1, \ldots, u_p) = \det(\langle u_i, u_j\rangle)_{1\le i,j\le p}$$

the Gram determinant [6] corresponding to the p-tuple $(u_1, \ldots, u_p)$. We also denote by $\nabla^2 f$ the hessian of f.

2.1. *Acceleration of* x_n. The following theorem shows that (h'_n) is well defined and converges faster than (x_n).

THEOREM 2.1. *Let $f : \mathbb{R}^p \to \mathbb{R}$ be of class C^2 and suppose that there exist constants $m > 0$ and $\varepsilon > 0$ such that*

$$m\|y\|^2 \le \langle \nabla^2 f(x)\cdot y, y\rangle, \qquad \forall (x, y) \in \mathbb{R}^p \times \mathbb{R}^p, \tag{2.1}$$

and for all $n \ge N$,

$$G\bigg(\frac{\nabla f(x_n)}{\|\nabla f(x_n)\|}, \frac{\nabla f(x_{n+1})}{\|\nabla f(x_{n+1})\|}, \ldots, \frac{\nabla f(x_{n+p-2})}{\|\nabla f(x_{n+p-2})\|}, \frac{\nabla f(x_{n+p-1})}{\|\nabla f(x_{n+p-1})\|}\bigg) \ge \varepsilon. \tag{2.2}$$

Then:

(1) *The sequence* $(h'_n)_n$ *is defined for* n *sufficiently large*,
(2) $\lim_{n\to\infty} \|h'_n - x^*\|/\|x_n - x^*\| = 0$.

The proof of this theorem is based on the following lemmas.

LEMMA 2.1. *Under the same assumptions as in Theorem* 2.1 *we have*

(1) ΔX_n *is regular for* n *sufficiently large*,
(2) $\|\Delta F'(x_n)\Delta X_n^{-1} - \nabla^2 f(x_n)\| \to 0$.

Proof. (1) First let us show that for all k, $\lambda_k \neq 0$ if $x_k \neq x^*$. Assume that there exists k_0 such that $\lambda_{k_0} = 0$. Then using (1.2) and (1.3) we get $x_{k_0+1} = x_{k_0}$ and $\langle \nabla f(x_{k_0+1}), \nabla f(x_{k_0})\rangle = 0$, i.e. $\|\nabla f(x_{k_0})\| = 0$, which gives $x_{k_0} = x^*$.

Now, we have

$$\Delta X_n = (x_{n+1} - x_n, \ldots, x_{n+p} - x_{n+p-1}),$$

and hence, by (1.2),

$$\Delta X_n = -(\lambda_n \nabla f(x_n), \ldots, \lambda_{n+p-1}\nabla f(x_{n+p-1})),$$

so as $\lambda_k \neq 0$ for all k it follows that ΔX_n is regular if and only if

$$\det(\nabla f(x_n), \ldots, \nabla f(x_{n+p-1})) \neq 0.$$

Thus by the hypothesis (2.2), ΔX_n is regular for $n \geq N$.

(2) We have

$$(\Delta F'(x_n)\Delta X_n^{-1} - \nabla^2 f(x_n))\Delta X_n = \Delta F'(x_n) - \nabla^2 f(x_n)\Delta X_n.$$

Thus, for every $i \in \{0, 1, \ldots, p-1\}$,

$$\begin{aligned}(\Delta F'(x_n)\Delta X_n^{-1} &- \nabla^2 f(x_n))\Delta x_{n+i} \\ &= \nabla f(x_{n+i+1}) - \nabla f(x_{n+i}) - \nabla^2 f(x_n)\Delta x_{n+i}.\end{aligned}$$

Applying the mean value theorem to ∇f (see [8]), we have

$$\begin{aligned}(\Delta F'(x_n)\Delta X_n^{-1} &- \nabla^2 f(x_n))\Delta x_{n+i} \\ &= \int_0^1 \nabla^2 f(x_{n+i} + t\Delta x_{n+i})\Delta x_{n+i}\,dt - \nabla^2 f(x_n)\Delta x_{n+i} \\ &= \int_0^1 (\nabla^2 f(x_{n+i} + t\Delta x_{n+i}) - \nabla^2 f(x_{n+i}))\Delta x_{n+i}\,dt \\ &\quad + (\nabla^2 f(x_{n+i}) - \nabla^2 f(x_n))\Delta x_{n+i}.\end{aligned}$$

Thus

$$\left\|(\Delta F'(x_n)\Delta X_n^{-1} - \nabla^2 f(x_n))\frac{\Delta x_{n+i}}{\|\Delta x_{n+i}\|}\right\|$$
$$\le \sup_{0\le t\le 1} \|\nabla^2 f(x_{n+i} + t\Delta x_{n+i}) - \nabla^2 f(x_{n+i})\| + \|\nabla^2 f(x_{n+i}) - \nabla^2 f(x_n)\|.$$

Since $\nabla^2 f$ is continuous, it follows that

$$\left\|(\Delta F'(x_n)\Delta X_n^{-1} - \nabla^2 f(x_n))\frac{\Delta x_{n+i}}{\|\Delta x_{n+i}\|}\right\| \to 0. \tag{2.3}$$

Let $a_n \in \mathbb{R}^p$ be such that $\|a_n\| = 1$ and

$$\|\Delta F'(x_n)\Delta X_n^{-1} - \nabla^2 f(x_n)\| = \|(\Delta F'(x_n)\Delta X_n^{-1} - \nabla^2 f(x_n))a_n\|.$$

Since $\{\nabla f(x_n)/\|\nabla f(x_n)\|, \ldots, \nabla f(x_{n+p-1})/\|\nabla f(x_{n+p-1})\|\}$ is a basis for $\mathbb{R}^p$, we have

$$a_n = \sum_{i=0}^{p-1} \alpha_i^n \frac{\nabla f(x_{n+i})}{\|\nabla f(x_{n+i})\|} = -\sum_{i=0}^{p-1} \alpha_i^n \frac{\|\lambda_{n+i}\|}{\lambda_{n+i}} \cdot \frac{\Delta x_{n+i}}{\|\Delta x_{n+i}\|}$$

and hence

$$\|\Delta F'(x_n)\Delta X_n^{-1} - \nabla^2 f(x_n)\| \le \sum_{i=0}^{p-1} |\alpha_i^n| \left\|(\Delta F'(x_n)\Delta X_n^{-1} - \nabla^2 f(x_n))\frac{\Delta x_{n+i}}{\|\Delta x_{n+i}\|}\right\|. \tag{2.4}$$

But, for all $j = 0, \ldots, p-1$,

$$\left\langle a_n, \frac{\nabla f(x_{n+j})}{\|\nabla f(x_{n+j})\|}\right\rangle = \sum_{i=0}^{p-1} \alpha_i^n \left\langle \frac{\nabla f(x_{n+i})}{\|\nabla f(x_{n+i})\|}, \frac{\nabla f(x_{n+j})}{\|\nabla f(x_{n+j})\|}\right\rangle,$$

thus $(\alpha_0^n, \ldots, \alpha_{p-1}^n)$ is a solution of a system whose determinant is the Gram determinant [6] corresponding to the p-tuple

$$\left(\frac{\nabla f(x_n)}{\|\nabla f(x_n)\|}, \ldots, \frac{\nabla f(x_{n+p-1})}{\|\nabla f(x_{n+p-1})\|}\right).$$

By Cramer's rule and (2.2),

$$\exists C > 0 : \quad |\alpha_i^n| \le C, \quad \forall n \ge N,\ \forall i = 0, \ldots, p-1,$$

thus, by (2.4),

$$\|\Delta F'(x_n)\Delta X_n^{-1} - \nabla^2 f(x_n)\| \le C \sum_{i=0}^{p-1} \left\|(\Delta F'(x_n)\Delta X_n^{-1} - \nabla^2 f(x_n))\frac{\Delta x_{n+i}}{\|\Delta x_{n+i}\|}\right\|.$$

Now (2.3) yields the assertion. ∎

Now we recall another lemma [6], which will be used in the proof of Theorem 2.1.

LEMMA 2.2. *Let A and B be two matrices. Assume that A is regular and that there exist α and β satisfying*

$$\|A^{-1}\| \le \alpha, \quad \|B - A\| \le \beta \quad \text{and} \quad \alpha\beta < 1.$$

Then

(1) *B is regular*,
(2) *$\|B^{-1}\| \le \alpha/(1-\alpha\beta)$.*

Proof of Theorem 2.1. (1) From Lemma 2.1 we have

$$\lim_{n\to\infty} \|\Delta F'(x_n)\Delta X_n^{-1} - \nabla^2 f(x_n)\| = 0.$$

$\nabla^2 f$ is continuous, therefore $\lim_{n\to\infty} \nabla^2 f(x_n) = \nabla^2 f(x^*)$ and

$$\lim_{n\to\infty} \|\Delta F'(x_n)\Delta X_n^{-1} - \nabla^2 f(x^*)\| = 0.$$

From (2.1), $\nabla^2 f(x^*)$ is regular; let $\alpha = \|\nabla^2 f(x^*)^{-1}\|$ and $\beta < 1/\alpha$. Then

$$\exists N > 0\ \forall n \ge N, \quad \|\Delta F'(x_n)\Delta X_n^{-1} - \nabla^2 f(x^*)\| \le \beta.$$

By Lemma 2.2, $\Delta F'(x_n)\Delta X_n^{-1}$ is regular and

$$\|(\Delta F'(x_n)\Delta X_n^{-1})^{-1}\| \le \frac{\alpha}{1-\alpha\beta}.$$

Therefore $\Delta F'(x_n)$ is also regular and

$$\|\Delta X_n(\Delta F'(x_n))^{-1}\| \le \frac{\alpha}{1-\alpha\beta}. \tag{2.5}$$

Thus there exists $N > 0$ such that for all $n \ge N$, the matrix $\Delta F'(x_n)$ is regular. Hence the first part of Theorem 2.1 is proved.

(2) Using (1.5) and the mean value theorem for ∇f, we have

$$\begin{aligned} h_n' - x^* &= \Delta X_n(\Delta F'(x_n))^{-1}(\Delta F'(x_n)\Delta X_n^{-1}(x_n - x^*) - \nabla f(x_n)) \\ &= \Delta X_n(\Delta F'(x_n))^{-1} \\ &\quad \times \Big(\Delta F'(x_n)\Delta X_n^{-1} - \int_0^1 \nabla^2 f(x^* + t(x_n - x^*))\,dt\Big)(x_n - x^*). \end{aligned}$$

By (2.5) we have

$$\begin{aligned} \|h_n' - x^*\| \le \frac{\alpha}{1-\alpha\beta}(&\|\Delta F'(x_n)\Delta X_n^{-1} - \nabla^2 f(x_n)\| \\ &+ \sup_{0\le t\le 1} \|\nabla^2 f(x^* + t(x_n - x^*)) - \nabla^2 f(x_n)\|) \cdot \|x_n - x^*\|. \end{aligned}$$

Using Lemma 2.1 completes the proof. ∎

Now we propose an algorithm for solving (1.1). Let us recall the Aitken Δ^2-process [3]. Let λ_0^0 be a positive scalar. We consider the sequence (λ_n^m) defined by

$$\lambda_n^{m+1} = \lambda_n^m + \alpha\langle\nabla f(x_n - \lambda_n^m \nabla f(x_n)), \nabla f(x_n)\rangle \tag{2.6}$$

for $n = 0, 1, \ldots$, and $m = 0, 1, \ldots$ The *Aitken Δ^2-process* is defined by

$$\Delta_2(\lambda_n^m) = \lambda_n^{m+1} - \frac{\Delta\lambda_n^{m+1}\Delta\lambda_n^m}{\Delta\lambda_n^{m+1} - \Delta\lambda_n^m} \tag{2.7}$$

where Δ is acting on the upper index. From (1.2), (1.3) and (1.5) we get the following algorithm.

Algorithm 1

a. Initialization: choose $x_0 \in \mathbb{R}^p, \alpha > 0, \varepsilon > 0$

b. Computation of $x_1, \ldots, x_p$

  for $n = 0, \ldots, p-1$, do

    b.1. choose $\lambda_n^0 > 0$

    b.2. compute λ_n^{m+1} by (2.6) for $m = 0, 1, \ldots$

    b.3. compute $\Delta_2(\lambda_n^m)$ by (2.7) for $m = 0, 1, \ldots$

      if $|\Delta_2(\lambda_n^{m+1}) - \Delta_2(\lambda_n^m)| \le \varepsilon$ then

        set $\lambda_n = \Delta_2(\lambda_n^m)$

        compute $x_{n+1} = x_n - \lambda_n \nabla f(x_n)$

      end if

  end do

c. Computation of h'_k

  set $h'_{-1} = x_p$

  for $k = 0, 1, \ldots$, do

    solve the linear system $\Delta F'(x_k)y_k = \nabla f(x_k)$

    compute $h'_k = x_k - \Delta X_k y_k$

    if $\|h'_k - h'_{k-1}\| \le \varepsilon$ then

      set $x^* = h'_k$

      stop

    end if

    compute x_{p+k+1} by b.1, b.2 and b.3

  end do.

Remarks 2.1. (1) The sequence $(\lambda_n^m)_m$ defined by (2.6) converges to the solution λ_n of (1.3) and we have (see [10])

$$\lim_{m\to\infty} \frac{\Delta_2(\lambda_n^m) - \lambda_n}{\lambda_n^m - \lambda_n} = 0.$$

(2) If $p = 2$ the hypothesis (2.2) is always satisfied because

$$G\left(\frac{\nabla f(x_n)}{\|\nabla f(x_n)\|}, \frac{\nabla f(x_{n+1})}{\|\nabla f(x_{n+1})\|}\right) = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}.$$

(3) If $p = 3$, we have

$$G\Big(\frac{\nabla f(x_n)}{\|\nabla f(x_n)\|}, \frac{\nabla f(x_{n+1})}{\|\nabla f(x_{n+1})\|}, \frac{\nabla f(x_{n+2})}{\|\nabla f(x_{n+2})\|}\Big) = \begin{vmatrix} 1 & 0 & \cos\theta^n_{n+2} \\ 0 & 1 & 0 \\ \cos\theta^n_{n+2} & 0 & 1 \end{vmatrix} = (\sin\theta^n_{n+2})^2,$$

where θ^n_{n+2} is the angle between $\nabla f(x_n)$ and $\nabla f(x_{n+2})$. We can see that this determinant can be zero, and (2.2) is satisfied if and only if

$$(\sin\theta^n_{n+2})^2 \geq \varepsilon.$$

(4) In general (2.2) is satisfied if and only if

$$\begin{vmatrix} 1 & 0 & \cos\theta^n_{n+2} & . & . & . & . & \cos\theta^n_{n+p-1} \\ 0 & 1 & 0 & \cos\theta^{n+1}_{n+3} & . & . & . & \cos\theta^{n+1}_{n+p-1} \\ \cos\theta^n_{n+2} & 0 & 1 & 0 & . & . & . & . \\ \cos\theta^n_{n+3} & \cos\theta^{n+1}_{n+3} & 0 & 1 & 0 & . & . & . \\ . & . & . & . & . & . & . & . \\ . & . & . & . & 0 & 1 & 0 & \cos\theta^{n+p-2}_{n+p-1} \\ . & . & . & . & . & 0 & 1 & 0 \\ \cos\theta^n_{n+p-1} & \cos\theta^{n+1}_{n+p-1} & . & . & . & \cos\theta^{n+p-2}_{n+p-1} & 0 & 1 \end{vmatrix} \geq \varepsilon,$$

where θ^{n+i}_{n+j} is the angle between $\nabla f(x_{n+i})$ and $\nabla f(x_{n+j})$ for $i \neq j$.

2.2. *Acceleration of* x_{n+p}. As we can see, the transformation h'_n is defined from $x_n, x_{n+1}, \ldots, x_{n+p}$. Hence it is more appropriate to compare $h'_n - x^*$ with $x_{n+p} - x^*$, instead of $x_n - x^*$. This comparison is based on the following theorem.

THEOREM 2.2. *Let* $f : \mathbb{R}^p \to \mathbb{R}$ *of class* C^2 *satisfy the conditions of Theorem* 2.1. *Then*

(2.8) $$h'_n = x_{n+p} - \Delta X_n(\Delta F'(x_n))^{-1}\nabla f(x_{n+p}), \quad n = 0, 1, \ldots$$

Proof. By (1.5) we have

$$h'_n = x_n - \Delta X_n(\Delta F'(x_n))^{-1}\nabla f(x_n), \quad n = 0, 1, \ldots$$

By the definition of $\Delta F'(x_n)$ we can easily see that

$$(\Delta F'(x_n))^{-1}\Delta\nabla f(x_{n+i}) = e_{i+1}, \quad i = 0, \ldots, p-1,$$

where e_j is the jth vector of the canonical basis of $\mathbb{R}^p$. We also have

$$\nabla f(x_n) = -\sum_{i=0}^{p-1}\Delta\nabla f(x_{n+i}) + \nabla f(x_{n+p}).$$

Thus

$$(\Delta F'(x_n))^{-1}\nabla f(x_n) = -\sum_{i=0}^{p-1} e_{i+1} + (\Delta F'(x_n))^{-1}\nabla f(x_{n+p}) = -e + (\Delta F'(x_n))^{-1}\nabla f(x_{n+p})$$

where $e = (1, 1, \ldots, 1)^t$, hence

$$h'_n = x_n + \Delta X_n e - \Delta X_n (\Delta F'(x_n))^{-1} \nabla f(x_{n+p});$$

but $\Delta X_n e = \sum_{i=0}^{p-1} \Delta x_{n+i} = -x_n + x_{n+p}$, hence

$$h'_n = x_{n+p} - \Delta X_n (\Delta F'(x_n))^{-1} \nabla f(x_{n+p}). \ \blacksquare$$

Remark 2.2. In general, from the relation

$$\nabla f(x_n) = -\sum_{i=0}^{k-1} \Delta \nabla f(x_{n+i}) + \nabla f(x_{n+k}),$$

we can easily see that

$$h'_n = x_{n+k} - \Delta X_n (\Delta F'(x_n))^{-1} \nabla f(x_{n+k}), \qquad k = 0, 1, \ldots, p.$$

Now, using the new form (2.8) of h'_n we can show the following result.

Theorem 2.3. *Under the same assumptions as in Theorem* 2.1, *we have*

$$\lim_{n\to\infty} \frac{\|h'_n - x^*\|}{\|x_{n+p} - x^*\|} = 0.$$

Proof. Using (2.8) and the mean value theorem for ∇f we have

$$\begin{aligned} h'_n - x^* &= \Delta X_n (\Delta F'(x_n))^{-1} (\Delta F'(x_n) \Delta X_n^{-1} (x_{n+p} - x^*) - \nabla f(x_{n+p})) \\ &= \Delta X_n (\Delta F'(x_n))^{-1} \\ &\quad \times \Big(\Delta F'(x_n) \Delta X_n^{-1} - \int_0^1 \nabla^2 f(x^* + t(x_{n+p} - x^*))\, dt\Big)(x_{n+p} - x^*). \end{aligned}$$

By (2.5), we obtain

$$\begin{aligned} \|h'_n - x^*\| \le \frac{\alpha}{1-\alpha\beta} (&\|\Delta F'(x_n) \Delta X_n^{-1} - \nabla^2 f(x_{n+p})\| \\ &+ \nabla^2 f(x^* + t_n(x_{n+p} - x^*)) - \nabla^2 f(x_{n+p})\|)\|x_{n+p} - x^*\| \end{aligned}$$

with $t_n \in [0,1]$. An application of Lemma 2.1 completes the proof. ■

Remark 2.3. From Remark 2.2, we also have

$$\lim_{n\to\infty} \frac{\|h'_n - x^*\|}{\|x_{n+k} - x^*\|} = 0, \qquad \forall k = 0, \ldots, p.$$

2.3. *Relation between h'_{n+1} and h'_n.* Now we study the relation between h'_{n+1} and h'_n. This is given in the following theorem.

Theorem 2.4. *Under the same assumptions as in Theorem* 2.1, *we have*

$$h'_{n+1} = h'_n - E_n \nabla f(x_n) \tag{2.9}$$

with

$$E_n = \Delta X_{n+1} (\Delta F'(x_{n+1}))^{-1} (I - \lambda_n A_n) + \lambda_n I - \Delta X_n (\Delta F'(x_n))^{-1}$$

and

$$A_n = \int_0^1 \nabla^2 f(x_n + t(x_{n+1} - x_n))\, dt.$$

Proof. By (1.2) and (1.5) we have

$$\begin{aligned}
h'_{n+1} - h'_n &= x_{n+1} - x_n \\
&\quad - (\Delta X_{n+1}(\Delta F'(x_{n+1}))^{-1}\nabla f(x_{n+1}) - \Delta X_n(\Delta F'(x_n))^{-1}\nabla f(x_n)) \\
&= -\lambda_n \nabla f(x_n) \\
&\quad - (\Delta X_{n+1}(\Delta F'(x_{n+1}))^{-1}\nabla f(x_{n+1}) - \Delta X_n(\Delta F'(x_n))^{-1}\nabla f(x_n)) \\
&= -(\Delta X_{n+1}(\Delta F'(x_{n+1}))^{-1}\nabla f(x_{n+1}) \\
&\quad + (\lambda_n I - \Delta X_n(\Delta F'(x_n))^{-1})\nabla f(x_n)).
\end{aligned}$$

Applying the mean value theorem to ∇f, we have

$$\begin{aligned}
\nabla f(x_{n+1}) &= \nabla f(x_n) + \int_0^1 \nabla^2 f(x_n + t(x_{n+1} - x_n))(x_{n+1} - x_n)\, dt \\
&= (I - \lambda_n A_n)\nabla f(x_n),
\end{aligned}$$

where $A_n = \int_0^1 \nabla^2 f(x_n + t(x_{n+1} - x_n))\, dt$. Then

$$\begin{aligned}
h'_{n+1} = h'_n &- (\Delta X_{n+1}(\Delta F'(x_{n+1}))^{-1}(I - \lambda_n A_n) \\
&+ (\lambda_n I - \Delta X_n(\Delta F'(x_n))^{-1}))\nabla f(x_n),
\end{aligned}$$

which is the required assertion. ■

Remark 2.4. Under certain assumptions we have

$$h'_{n+1} = h'_n - D_n^{-1}\nabla f(h'_n).$$

Indeed, (1.5) and the mean value theorem imply

$$\begin{aligned}
\nabla f(h'_n) &= \nabla f(x_n) + \int_0^1 \nabla^2 f(x_n + t(h'_n - x_n))(h'_n - x_n)\, dt \\
&= (I - B_n C_n)\nabla f(x_n)
\end{aligned}$$

where

$$B_n = \int_0^1 \nabla^2 f(x_n + t(h'_n - x_n))\, dt, \qquad C_n = \Delta X_n(\Delta F'(x_n))^{-1}.$$

If $I - B_n C_n$ is regular, then by (2.9) we have

$$h'_{n+1} = h'_n - E_n(I - B_n C_n)^{-1}\nabla f(h'_n).$$

If also E_n is regular, we have

$$h'_{n+1} = h'_n - D_n^{-1}\nabla f(h'_n)$$

with $D_n = (I - B_n C_n)E_n^{-1}$.

3. Numerical experiments. In this section, we present some numerical experiments. We compare the modified Henrici transformation MHT given by Algorithm 1 with the gradient method with optimal step GMO [1, 4]. This comparison will be summarized in tables which give the number of iterations, iter, and the associated residual norms for each method. The stopping criterion is given by res $= \|x_k - x^*\|$, where x^* is the solution of problem (1.1). To solve the linear system in Algorithm 1 we use Gaussian elimination.

EXAMPLE 1. The first example has been used in [5]. We consider the function

$$f(x_1, x_2) = \tfrac{1}{2}(x_1)^2 + \tfrac{9}{2}x_2^2.$$

We find that $x^* = (0,0)$. We take $x_0 = (9,1)$. The results are summarized in Table 1.

TABLE 1

iter	GMO	MTH
0	9.055385138137417E-000	9.055385138137417E-000
1	7.244308110509934E-000	1.110223024625157E-016
21	8.352118606604707E-002	
100	1.844614530595065E-009	
155	8.626902901469026E-015	

EXAMPLE 2. This example is taken from [4, p. 194]. We consider the function

$$f(x,y) = \tfrac{1}{2}(\alpha x^2 + \beta y^2)$$

with $\alpha = 1/2$ and $\beta = 1$. We find that $x^* = (0,0)$. For different initial guess points x_0 we obtain Tables 2.1–2.4.

TABLE 2.1. $x_0 = (2,1)$

iter	GMO	MTH
0	2.236067977499790E-000	2.236067977499790E-000
1	7.453559924999299E-001	2.220446049250313E-016
3	8.281733249999221E-002	
15	1.558354219941484E-007	
31	3.620146166165558E-015	

TABLE 2.2. $x_0 = (1, 0.5)$

iter	GMO	MTH
0	2.236067977499790E-000	2.236067977499790E-000
1	7.453559924999299E-001	2.220446049250313E-016
3	8.281733249999221E-002	
15	1.558354219941484E-007	
31	3.620146166165558E-015	

TABLE 2.3. $x_0 = (1, 0.1)$

iter	GMO	MTH
0	1.004987562112089E-000	1.004987562112089E-000
1	9.972527420619452E-002	2.775557561562891E-017
3	1.810553271717403E-003	
10	1.982398812895594E-009	
17	1.177209576653762E-015	

TABLE 2.4. $x_0 = (20, 10)$

iter	GMO	MTH
0	22.360679774997900E-000	22.360679774997900E-000
1	7.453559924999300E-000	1.776356839400251E-015
5	9.201925833332468E-002	
15	1.558354219941484E-006	
33	4.022384629072841E-015	

EXAMPLE 3. This example is taken from [5]. We consider the function

$$f(x, y) = (xy + 1)^2 + (y + 1)^2.$$

We find that $x^* = (1, -1)$. We obtain the following results for different x_0.

TABLE 3.1. $x_0 = (0, 1)$

iter	GMO	MTH
0	2.236067977499790E-000	2.236067977499790E-000
1	1.788854381999832E-000	4.831747651906473E-001
20	1.074461082435433E-004	1.724424625459104E-008
39	6.286983084653295E-008	4.848786246276754E-015
60	1.679690792659464E-011	
72	9.943763361765949E-015	

TABLE 3.2. $x_0 = (0.1, 1)$

iter	GMO	MTH
0	2.193171219946131E-000	1.118033988749895E-000
1	1.722618673223975E-000	4.614411326269345E-001
10	8.705111891125794E-004	5.384368811483133E-007
23	5.816252278905881E-008	2.809445799405440E-015
30	3.285322337075431E-010	
45	4.785995741689946E-015	

TABLE 3.3. $x_0 = (-3, 3)$

iter	GMO	MTH
0	5.656854249492381E-000	5.656854249492381E-000
1	4.338609156373126E-001	3.551097168625910E-002
5	2.711170911269893E-004	1.161126535039719E-007
10	3.430602253738676E-008	1.471934093453127E-015
15	4.339406371833765E-012	
19	3.338112220702277E-015	

TABLE 3.4. $x_0 = (1.01, -1.01)$

iter	GMO	MTH
0	1.414213562373096E-000	1.414213562373096E-000
1	2.741752656369141E-003	1.222262718678914E-005
5	5.573590389105667E-006	4.475068539078773E-011
8	5.350830654762778E-008	3.054723882208400E-015
15	1.048229607808129E-012	
19	2.373570351140066E-015	

Example 4. This example is taken from [5]. We consider the function

$$f(x) = f_1^2(x) + f_2^2(x),$$

where $x = (x_1, x_2)$, $f_1(x) = x_1^2 - 2x_2 + 3$ and $f_2(x) = x_1 x_2 - 2$. We find that $x^* = (1, 2)$. For different initial points x_0 we obtain the following results.

TABLE 4.1. $x_0 = (1.5, 1.5)$

iter	GMO	MTH
0	7.071067811865476E-001	7.071067811865476E-001
1	1.821950074697108E-001	5.936182561260886E-004
2	2.118892238035503E-003	7.812323171249536E-006
3	8.628089765268629E-005	4.425839343824471E-009
4	3.444166556862442E-006	5.437020644362115E-012
6	5.491799555959319E-009	1.374886346414612E-017
9	3.496437483938589E-013	
11	5.334779303576802E-016	

TABLE 4.2. $x_0 = (0, 0)$

iter	GMO	MTH
0	2.236067977499790E-000	2.236067977499790E-000
1	1.000000000000000E-000	4.110657724161396E-001
2	4.472135954999579E-001	9.872713013407693E-002
5	5.960181541586701E-004	1.350968601863407E-007
7	3.472741049050291E-006	4.639123236287278E-012
9	2.017253030954468E-008	1.565837193626824E-016
11	1.171765719968503E-010	
13	6.806166568826876E-013	
15	4.092329968692928E-015	

TABLE 4.3. $x_0 = (-1, 0)$

iter	GMO	MTH
0	2.828427124746190E-000	2.828427124746190E-000
1	3.999999999999999E-001	2.346678418411335E-002
2	1.154829939066024E-001	2.555952658965605E-003
4	3.535120882045415E-003	5.075870513349970E-006
6	1.125058682735095E-004	5.373992891221222E-009
8	3.584010172629917E-006	5.461683032157274E-012
10	1.141764851087469E-007	5.543210949557351E-015
12	3.637345311678336E-009	
15	2.068216140766206E-011	
20	3.821435084792773E-015	

TABLE 4.4. $x_0 = (1.4, 1.6)$

iter	GMO	MTH
0	5.656854249492379E-001	5.656854249492379E-001
1	1.374420072475986E-001	1.476869385644176E-003
2	6.448644012207421E-003	3.107703768385792E-005
4	2.982170657758973E-005	4.235319264903191E-010
6	1.333807714493249E-007	8.461361619985638E-015
8	5.964656961345872E-010	
10	2.667220740432435E-012	
13	8.437427167243345E-016	

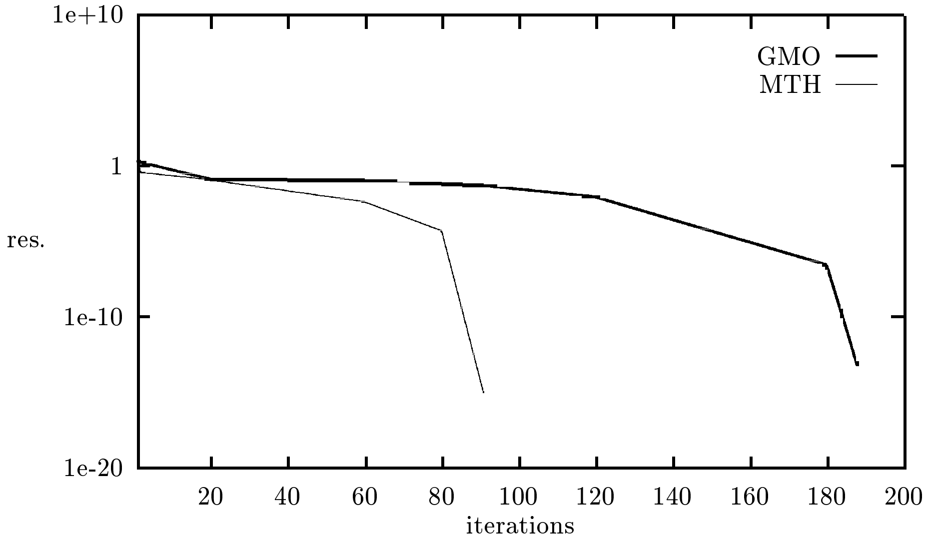

Fig. 1. Example 5 with $n = 2$, initial point $= (-1, 2)$

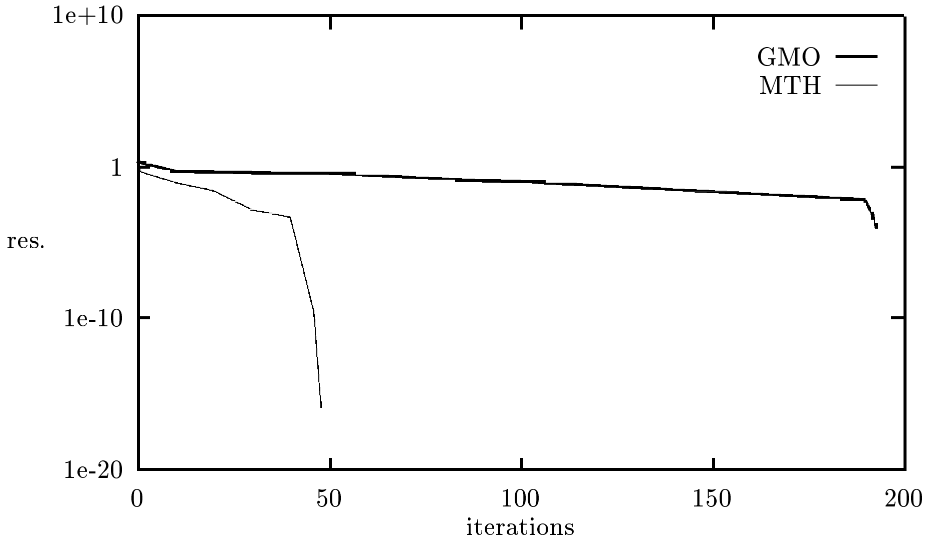

Fig. 2. Example 5 with $n = 4$, initial point $= (-1, 2, 0.8, 0.9)$

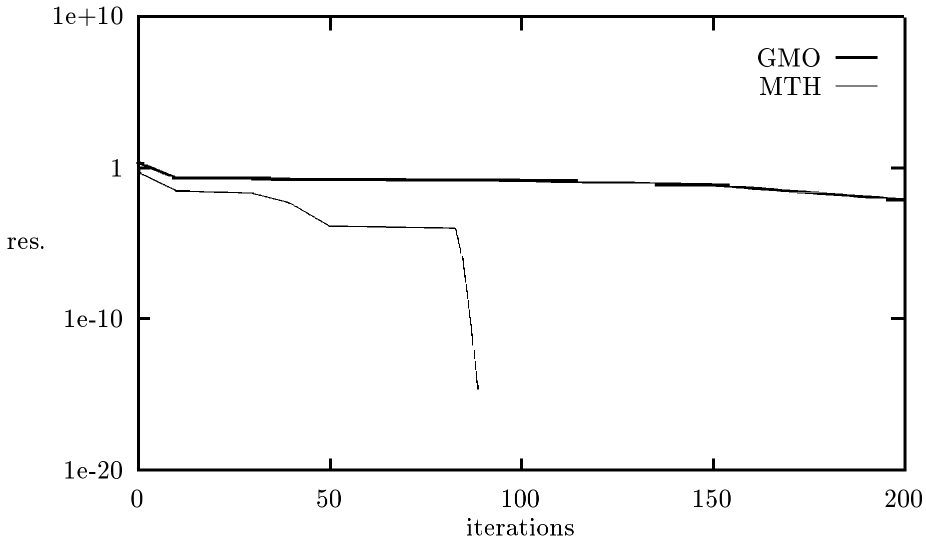

Fig. 3. Example 5 with $n = 10$, initial point $= (4, -1, 3, 2, 5, 0.8, 0.5, 0.7, 1, 0.5)$

Example 5. This example is taken from [5]. We consider the function

$$f(x) = \sum_{1\le i\le n} f_i^2(x),$$

where n is any positive multiple of 2, $x = (x_i)_{1\le i\le n}$, and for $i = 1, \ldots, n/2$, $f_{2i-1}(x) = 10(x_{2i} - x_{2i-1}^2)$, $f_{2i}(x) = 1 - x_{2i-1}$. We find that $x^* = (1, \ldots, 1)$.

We use three values of n: $n = 2$ (Fig. 1), $n = 4$ (Fig. 2) and $n = 10$ (Fig. 3).

Throughout these examples we can see that the modified Henrici transformation MHT given by Algorithm 1 converges faster than the gradient method with optimal step GMO.

References

[1] A. Auslender, *Optimisation: Méthodes Numériques*, Masson, Paris, 1976.
[2] C. Brezinski, *Accélération de la Convergence en Analyse Numérique*, Springer, Berlin, 1977.
[3] —, *Algorithmes d'Accélération de la Convergence*, *Etude Numérique*, Editions Technip, Paris, 1978.
[4] P. G. Ciarlet, *Introduction à l'Analyse Numérique Matricielle et à l'Optimisation*, Masson, Paris, 1985.
[5] J. E. Dennis, Jr. and R. B. Schnabel, *Numerical Methods For Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, Englewood Cliffs, NJ, 1983.
[6] I. Gohberg and S. Golderg, *Basic Operator Theory*, Birkhäuser, Basel, 1981.
[7] P. Henrici, *Elements of Numerical Analysis*, Wiley, New York, 1964.
[8] A. Kolmogorov et S. Fomine, *Eléments de la Théorie des Fonctions et de l'Analyse Fonctionnelle*, Editions Mir, Moscou, 1979.
[9] H. Le Ferrand, *Recherche d'extrema par les méthodes d'extrapolation*, C. R. Acad. Sci. Paris Sér. I 318 (1994), 1043–1049.
[10] B. Rhanizar, *On extrapolation methods in optimization*, Appl. Numer. Math. 25 (1997), 485–498.
[11] H. Sadok, *Accéleration de la convergence de suites vectorielles et méthodes point fixe*, Thèse, Université des Sciences et Techniques de Lille, 1988.
[12] J. Vignes, *Algorithmes Numériques*, *Analyse et Mise en Œuvre*, Vol. 2, Editions Technip, Paris, 1980.

Département de Mathématiques
Ecole Normale Supérieure Takaddoum
B.P. 5118
Rabat, Maroc

Département de Mathématiques
Faculté des Sciences Semlalia
Université Cadi Ayyad
Marrakech, Maroc

Received on 20.3.1998