ON HENRICI’S TRANSFORMATION IN OPTIMIZATION

Abstract. Henrici’s transformation is a generalization of Aitken’s $\Delta^2$-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 $\Delta^2$-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 $\|x\| = (\sum_{1 \leq i \leq p} x_i^2)^{1/2}$. We denote by $\langle \cdot, \cdot \rangle$ 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:

\begin{align}
(1.1) \quad & \text{find } x^* \in \mathbb{R}^p \text{ such that } f(x^*) = \min_{x \in \mathbb{R}^p} f(x), \\
(1.2) \quad & x_{n+1} = x_n - \lambda_n \nabla f(x_n), \quad n = 0, 1, \ldots, \\
(1.3) \quad & f(x_n - \lambda_n \nabla f(x_n)) = \min_{\lambda \in \mathbb{R}} f(x_n - \lambda \nabla f(x_n)),
\end{align}

where $\nabla f$ is the gradient of $f$. Under certain assumptions, the sequence
\((x_n)\) 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

\[
(1.4) \quad h_n = x_n - \Delta X_n (\Delta^2 X_n)^{-1} \Delta x_n, \quad n = 0, 1, \ldots,
\]

where \(\Delta X_n\) 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:

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

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 convergence acceleration results. Consider the sequence \((x_n)\) 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

\[
G(u_1, \ldots, u_p) = \det((u_i, u_j))_{1 \leq i, j \leq p}
\]

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. Convergence acceleration results.** Consider the sequence \((x_n)\) 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

\[
G(u_1, \ldots, u_p) = \det((u_i, u_j))_{1 \leq i, j \leq p}
\]

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

\[
(2.1) \quad m\|y\|^2 \leq \langle \nabla^2 f(x) \cdot y, y \rangle, \quad \forall (x, y) \in \mathbb{R}^p \times \mathbb{R}^p,
\]

and for all \(n \geq N\),

\[
(2.2) \quad G\left(\frac{\nabla f(x_n)}{\|\nabla f(x_n)\|}, \frac{\nabla f(x_{n+1})}{\|\nabla f(x_{n+1})\|}, \ldots, \frac{\nabla f(x_{n+p-2})}{\|\nabla f(x_{n+p-2})\|}, \frac{\nabla f(x_{n+p-1})}{\|\nabla f(x_{n+p-1})\|}\right) \geq \varepsilon.
\]
Then:

(1) The sequence \((h'_n)\) 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) \(\Delta X_n\) 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

\[ x_{k_0 + 1} = x_{k_0} \quad \text{and} \quad \langle \nabla f(x_{k_0} + 1), \nabla f(x_{k_0}) \rangle = 0, \]

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 \(\Delta X_n\) 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), \(\Delta X_n\) 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 \(i \in \{0, 1, \ldots, p - 1\}\),

\[ (\Delta F'(x_n)\Delta X_n^{-1} - \nabla^2 f(x_n))\Delta x_{n+i} = \nabla f(x_{n+i+1}) - \nabla f(x_{n+i}) - \nabla^2 f(x_n)\Delta x_{n+i}. \]

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

\[ (\Delta F'(x_n)\Delta X_n^{-1} - \nabla^2 f(x_n))\Delta x_{n+i} \]

\[ = \int_0^1 \nabla^2 f(x_{n+i} + t\Delta x_{n+i}) \Delta x_{n+i} \, dt - \nabla^2 f(x_n)\Delta x_{n+i} \]

\[ = \int_0^1 (\nabla^2 f(x_{n+i} + t\Delta x_{n+i}) - \nabla^2 f(x_{n+i})) \Delta x_{n+i} \, dt \]

\[ + (\nabla^2 f(x_{n+i}) - \nabla^2 f(x_n))\Delta x_{n+i}. \]
Thus
\[ \left\| \nabla f(x_n) \Delta X_n^{-1} - \nabla^2 f(x_n) \right\| \frac{\Delta x_{n+i}}{\| \Delta x_{n+i} \|} \leq \sup_{0 \leq t \leq 1} \| \nabla^2 f(x_{n+i} + t\Delta x_{n+i}) - \nabla^2 f(x_{n+i}) \| + \| \nabla^2 f(x_{n+i}) - \nabla^2 f(x_n) \|. \]

Since \( \nabla^2 f \) is continuous, it follows that
\[ (2.3) \quad \left\| (\nabla f(x_n) \Delta X_n^{-1} - \nabla^2 f(x_n)) \right\| \frac{\Delta x_{n+i}}{\| \Delta x_{n+i} \|} \rightarrow 0. \]

Let \( a_n \in \mathbb{R}^p \) be such that \( \| a_n \| = 1 \) and
\[ \| \nabla f(x_n) \Delta X_n^{-1} - \nabla^2 f(x_n) \| = \| (\nabla f(x_n) \Delta X_n^{-1} - \nabla^2 f(x_n))a_n \|. \]
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
\[ a_n = \sum_{i=0}^{p-1} \alpha_i^n \nabla f(x_{n+i}) = \sum_{i=0}^{p-1} \alpha_i^n \lambda_{n+i} \frac{\Delta x_{n+i}}{\| \Delta x_{n+i} \|} \]
and hence
\[ (2.4) \quad \| \nabla f(x_n) \Delta X_n^{-1} - \nabla^2 f(x_n) \| \leq \sum_{i=0}^{p-1} |\alpha_i^n| \left\| (\nabla f(x_n) \Delta X_n^{-1} - \nabla^2 f(x_n)) \right\| \frac{\Delta x_{n+i}}{\| \Delta x_{n+i} \|}. \]

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),
\[ \exists C > 0 : \quad |\alpha_i^n| \leq C, \quad \forall n \geq N, \quad \forall i = 0, \ldots, p-1, \]
thus, by (2.4),
\[ \| \nabla f(x_n) \Delta X_n^{-1} - \nabla^2 f(x_n) \| \leq C \sum_{i=0}^{p-1} \left\| (\nabla f(x_n) \Delta X_n^{-1} - \nabla^2 f(x_n)) \right\| \frac{\Delta x_{n+i}}{\| \Delta x_{n+i} \|}. \]

Now (2.3) yields the assertion. \( \blacksquare \)
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 $\alpha$ and $\beta$ satisfying
\[
\|A^{-1}\| \leq \alpha, \quad \|B - A\| \leq \beta \quad \text{and} \quad \alpha \beta < 1.
\]

Then

1. $B$ is regular,
2. $\|B^{-1}\| \leq \frac{\alpha}{1 - \alpha \beta}$.

**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
\[
\exists N > 0 \forall n \geq N, \quad \|\Delta F'(x_n) \Delta X_n^{-1} - \nabla^2 f(x^*)\| \leq \beta.
\]
By Lemma 2.2, $\Delta F'(x_n) \Delta X_n^{-1}$ is regular and
\[
\|\Delta F'(x_n) \Delta X_n^{-1}\| \leq \frac{\alpha}{1 - \alpha \beta}.
\]
Therefore $\Delta F'(x_n)$ is also regular and
\[(2.5) \quad \|\Delta X_n (\Delta F'(x_n))^{-1}\| \leq \frac{\alpha}{1 - \alpha \beta}.
\]
Thus there exists $N > 0$ such that for all $n \geq N$, 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 $\nabla f$, we have
\[
h'_n - x^* = \Delta X_n (\Delta F'(x_n))^{-1} (\Delta F'(x_n) \Delta X_n^{-1} (x_n - x^*) - \nabla f(x_n))
\]
\[
= \Delta X_n (\Delta F'(x_n))^{-1} \times \left( \Delta F'(x_n) \Delta X_n^{-1} \int_0^1 \nabla^2 f(x^* + t(x_n - x^*)) \, dt \right) (x_n - x^*).
\]
By (2.5) we have
\[
\|h'_n - x^*\| \leq \frac{\alpha}{1 - \alpha \beta} \|\Delta F'(x_n) \Delta X_n^{-1} - \nabla^2 f(x_n)\|
\]
\[
+ \sup_{0 \leq t \leq 1} \|\nabla^2 f(x^* + t(x_n - x^*)) - \nabla^2 f(x_n)\| \cdot \|x_n - x^*\|.
\]
Using Lemma 2.1 completes the proof. \[\blacksquare\]
Now we propose an algorithm for solving (1.1). Let us recall the Aitken $\Delta^2$-process [3]. Let $\lambda_0^0$ be a positive scalar. We consider the sequence $(\lambda_m^n)$ defined by

$$(2.6) \quad \lambda_{m+1}^n = \lambda_m^n + \alpha (\nabla f(x_n - \lambda_m^n \nabla f(x_n)), \nabla f(x_n))$$

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

$$(2.7) \quad \Delta_2(\lambda_m^n) = \lambda_{m+1}^n - \frac{\Delta \lambda_{m+1}^{m+1} \Delta \lambda_m^m}{\Delta \lambda_{m+1}^m - \Delta \lambda_m^m}$$

where $\Delta$ 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 $\lambda_0^n > 0$

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

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

      if $|\Delta_2(\lambda_{m+1}^n) - \Delta_2(\lambda_m^n)| \leq \varepsilon$ then

      set $\lambda_n = \Delta_2(\lambda_m^n)$

      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 $\|h'_k - h'_{k-1}\| \leq \varepsilon$ 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 $(\lambda_m^n)_m$ defined by (2.6) converges to the solution $\lambda_n$ of (1.3) and we have (see [10])

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

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

$$G = \begin{vmatrix} \nabla f(x_n) & \nabla f(x_{n+1}) \\ \|\nabla f(x_n)\| & \|\nabla f(x_{n+1})\| \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}.$$
Thus
\[ \Delta F \]
By the definition of
\[ \theta \]
where \( \theta_{n+2} \) 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
\[ (\sin \theta_{n+2})^2 \geq \varepsilon. \]

(4) In general (2.2) is satisfied if and only if
\[
\begin{vmatrix}
1 & 0 & \cos \theta_{n+2} & \cdots & \cdots & \cos \theta_{n+p-1} \\
0 & 1 & 0 & \cos \theta_{n+3} & \cdots & \cos \theta_{n+p-1} \\
\cos \theta_{n+2} & 0 & 1 & 0 & \cdots & \cdots \\
\cos \theta_{n+3} & \cos \theta_{n+3} & 0 & 1 & 0 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\cos \theta_{n+p-1} & \cos \theta_{n+p-1} & \cdots & \cdots & \cos \theta_{n+p-2} & 0 & 1
\end{vmatrix} \geq \varepsilon,
\]
where \( \theta_{n+j}^{m+i} \) 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
\[
(2.8) \quad h_n' = x_{n+p} - \Delta X_n(\Delta F'(x_n))^{-1}\nabla f(x_{n+p}), \quad n = 0, 1, \ldots
\]

**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 \( j \)th 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
\[
h'_n = x_{n+p} - \Delta X_n (\Delta F'(x_n))^{-1} \nabla f(x_{n+p}). \quad \Box
\]

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

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 $\nabla f$ we have
\[
h'_n - x^* = \Delta X_n (\Delta F'(x_n))^{-1} (\Delta F'(x_n) \Delta X_n^{-1} (x_{n+p} - x^*) - \nabla f(x_{n+p}))
\]
\[
= \Delta X_n (\Delta F'(x_n))^{-1}
\]
\[
\times \left( \Delta F'(x_n) \Delta X_n^{-1} - \frac{1}{0} \nabla^2 f(x^* + t(x_{n+p} - x^*)) dt \right) (x_{n+p} - x^*).
\]

By (2.5), we obtain
\[
\|h'_n - x^*\| \leq \frac{\alpha}{1 - \alpha \beta} (\|\Delta F'(x_n) \Delta X_n^{-1} - \nabla^2 f(x_{n+p})\|)
\]
\[
+ \|\nabla^2 f(x^* + t_n(x_{n+p} - x^*)) - \nabla^2 f(x_{n+p})\| ||x_{n+p} - x^||
\]
with $t_n \in [0,1]$. An application of Lemma 2.1 completes the proof. \Box

**Remark 2.3.** From Remark 2.2, we also have
\[
\lim_{n \to \infty} \frac{\|h'_n - x^*\|}{\|x_{n+k} - x^*\|} = 0, \quad \forall k = 0, \ldots, p.
\]

**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
\[
(2.9) \quad h'_{n+1} = h'_n - E_n \nabla f(x_n)
\]
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
\[ h'_{n+1} - h'_n = x_{n+1} - x_n \]
\[ = \frac{\lambda_n}{\nabla f(x_n)} \]
\[ = -\lambda_n \nabla f(x_n) \]
\[ = -\lambda_n \nabla f(x_n) \]
\[ + (\lambda_n I - \Delta X_n \Delta F'(x_n)^{-1}) \nabla f(x_n). \]
Applying the mean value theorem to \( \nabla f \), we have
\[ \nabla f(x_{n+1}) = \nabla f(x_n) + \int_0^1 \nabla^2 f(x_n + t(x_{n+1} - x_n))(x_{n+1} - x_n) \, dt \]
where \( A_n = \int_0^1 \nabla^2 f(x_n + t(x_{n+1} - x_n)) \, dt. \) Then
\[ h'_{n+1} = h'_n - \Delta X_n \Delta F'(x_n)^{-1}(I - \lambda_n A_n) \]
\[ = (I - \lambda_n A_n) \nabla f(x_n), \]
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
\[ \nabla f(h'_n) = \nabla f(x_n) + \int_0^1 \nabla^2 f(x_n + t(h'_n - x_n))(h'_n - x_n) \, dt \]
\[ = (I - B_n C_n) \nabla f(x_n) \]
where
\[ B_n = \int_0^1 \nabla^2 f(x_n + t(h'_n - x_n)) \, dt, \quad C_n = \Delta X_n \Delta F'(x_n)^{-1}. \]
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

$$h_{n+1}^r = h_n^r - D_n^{-1} \nabla f(h_n^r)$$

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 $\text{res} = \|x_k - x^*\|$, 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

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

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

<table>
<thead>
<tr>
<th>iter</th>
<th>GMO</th>
<th>MTH</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>9.05385138137417E-000</td>
<td>9.05385138137417E-000</td>
</tr>
<tr>
<td>1</td>
<td>7.244308110509934E-000</td>
<td>1.110223024625157E-016</td>
</tr>
<tr>
<td>21</td>
<td>8.352118606604707E-002</td>
<td>8.626902901469026E-015</td>
</tr>
<tr>
<td>100</td>
<td>1.84461453095065E-009</td>
<td>8.626902901469026E-015</td>
</tr>
<tr>
<td>155</td>
<td>8.626902901469026E-015</td>
<td>8.626902901469026E-015</td>
</tr>
</tbody>
</table>

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

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

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>
<thead>
<tr>
<th>iter</th>
<th>GMO</th>
<th>MTH</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2.236067977499790E-000</td>
<td>2.236067977499790E-000</td>
</tr>
<tr>
<td>1</td>
<td>7.453559924999299E-001</td>
<td>2.220446049250313E-016</td>
</tr>
<tr>
<td>3</td>
<td>8.281733249999922E-002</td>
<td>2.220446049250313E-016</td>
</tr>
<tr>
<td>15</td>
<td>1.558354219941484E-007</td>
<td>2.220446049250313E-016</td>
</tr>
<tr>
<td>31</td>
<td>3.620146166165558E-015</td>
<td>2.220446049250313E-016</td>
</tr>
</tbody>
</table>
TABLE 2.2. $x_0 = (1, 0.5)$

<table>
<thead>
<tr>
<th>iter</th>
<th>GMO</th>
<th>MTH</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2.236067977499790E-000</td>
<td>2.236067977499790E-000</td>
</tr>
<tr>
<td>1</td>
<td>7.45355924999299E-001</td>
<td>2.220446049250313E-016</td>
</tr>
<tr>
<td>3</td>
<td>8.28173249999221E-002</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>1.558354219941484E-007</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>3.620146166165558E-015</td>
<td></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>iter</th>
<th>GMO</th>
<th>MTH</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.004987562112089E-000</td>
<td>1.004987562112089E-000</td>
</tr>
<tr>
<td>1</td>
<td>9.72527420619452E-002</td>
<td>2.775557561562891E-017</td>
</tr>
<tr>
<td>3</td>
<td>1.810553271717403E-003</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>1.982398812895594E-009</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>1.17720576653762E-015</td>
<td></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>iter</th>
<th>GMO</th>
<th>MTH</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>22.360679774997900E-000</td>
<td>22.360679774997900E-000</td>
</tr>
<tr>
<td>1</td>
<td>7.45355924999300E-000</td>
<td>1.776356839400251E-015</td>
</tr>
<tr>
<td>5</td>
<td>9.20192583332468E-002</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>8.705111891125794E-004</td>
<td>5.384368811483133E-007</td>
</tr>
<tr>
<td>33</td>
<td>4.022384629072841E-015</td>
<td></td>
</tr>
</tbody>
</table>

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. $x_0 = (0, 1)$

<table>
<thead>
<tr>
<th>iter</th>
<th>GMO</th>
<th>MTH</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2.236067977499790E-000</td>
<td>2.236067977499790E-000</td>
</tr>
<tr>
<td>1</td>
<td>1.788854381999832E-000</td>
<td>4.83174671906473E-001</td>
</tr>
<tr>
<td>20</td>
<td>1.074461082435433E-004</td>
<td>1.7244265459104E-008</td>
</tr>
<tr>
<td>39</td>
<td>6.286983084653295E-008</td>
<td>4.848786246276754E-015</td>
</tr>
<tr>
<td>60</td>
<td>1.67960792659464E-011</td>
<td></td>
</tr>
<tr>
<td>72</td>
<td>9.943763361765949E-015</td>
<td></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>iter</th>
<th>GMO</th>
<th>MTH</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2.193171219946131E-000</td>
<td>1.118033988749895E-000</td>
</tr>
<tr>
<td>1</td>
<td>1.72261867322975E-000</td>
<td>4.61441326269345E-001</td>
</tr>
<tr>
<td>10</td>
<td>8.705111891125794E-004</td>
<td>5.384368811483133E-007</td>
</tr>
<tr>
<td>23</td>
<td>5.81625227905881E-008</td>
<td>2.80945799405440E-015</td>
</tr>
<tr>
<td>30</td>
<td>3.28532237075431E-010</td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>4.78599574169946E-015</td>
<td></td>
</tr>
</tbody>
</table>
TABLE 3.3. \( x_0 = (-3,3) \)

<table>
<thead>
<tr>
<th>iter</th>
<th>GMO</th>
<th>MTH</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5.65854249492381E-000</td>
<td>5.65854249492381E-000</td>
</tr>
<tr>
<td>1</td>
<td>4.338609156373126E-001</td>
<td>3.551097168625910E-002</td>
</tr>
<tr>
<td>5</td>
<td>2.711170911269893E-004</td>
<td>1.161126535039719E-007</td>
</tr>
<tr>
<td>10</td>
<td>3.43062253738676E-008</td>
<td>1.47193493453127E-015</td>
</tr>
<tr>
<td>15</td>
<td>4.339406371833765E-012</td>
<td>4.33811222070227E-015</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>iter</th>
<th>GMO</th>
<th>MTH</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.41421356237096E-000</td>
<td>1.41421356237096E-000</td>
</tr>
<tr>
<td>1</td>
<td>2.74175265636941E-003</td>
<td>1.22286718678914E-005</td>
</tr>
<tr>
<td>5</td>
<td>5.57359038910566E-006</td>
<td>4.47506539078773E-011</td>
</tr>
<tr>
<td>8</td>
<td>5.35083065476277E-008</td>
<td>3.05472388208400E-015</td>
</tr>
<tr>
<td>15</td>
<td>1.048229607808129E-012</td>
<td>2.37357935114996E-015</td>
</tr>
</tbody>
</table>

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 \( f_2(x) = x_1x_2 - 2 \). We find that \( x^* = (1, 2) \). For different initial points \( x_0 \) we obtain the following results.

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

<table>
<thead>
<tr>
<th>iter</th>
<th>GMO</th>
<th>MTH</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>7.071067811865476E-001</td>
<td>7.071067811865476E-001</td>
</tr>
<tr>
<td>1</td>
<td>1.82195007469108E-001</td>
<td>5.93612561260886E-044</td>
</tr>
<tr>
<td>2</td>
<td>2.118892238035503E-003</td>
<td>7.8213211249536E-006</td>
</tr>
<tr>
<td>3</td>
<td>8.626897652686269E-005</td>
<td>4.4258394382447E-009</td>
</tr>
<tr>
<td>4</td>
<td>3.444166556862442E-006</td>
<td>5.43720644362115E-012</td>
</tr>
<tr>
<td>6</td>
<td>5.491799555959319E-009</td>
<td>1.37486346414612E-017</td>
</tr>
<tr>
<td>9</td>
<td>3.496437483938589E-013</td>
<td>2.37357935114996E-015</td>
</tr>
<tr>
<td>11</td>
<td>5.334779303576802E-016</td>
<td>5.334779303576802E-016</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>iter</th>
<th>GMO</th>
<th>MTH</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2.236067977499790E-000</td>
<td>2.236067977499790E-000</td>
</tr>
<tr>
<td>1</td>
<td>1.000000000000000E-000</td>
<td>4.110657724161396E-001</td>
</tr>
<tr>
<td>2</td>
<td>4.472135954999579E-001</td>
<td>9.872713013407936E-002</td>
</tr>
<tr>
<td>5</td>
<td>5.960181515586701E-004</td>
<td>1.35968601863407E-007</td>
</tr>
<tr>
<td>7</td>
<td>3.47271409650291E-006</td>
<td>4.63912326328727E-012</td>
</tr>
<tr>
<td>9</td>
<td>2.01725303954468E-008</td>
<td>1.56583719362682E-016</td>
</tr>
<tr>
<td>11</td>
<td>1.171765719968503E-010</td>
<td>4.09232996892928E-015</td>
</tr>
<tr>
<td>13</td>
<td>6.80616658828687E-013</td>
<td>8.80616658828687E-013</td>
</tr>
<tr>
<td>15</td>
<td>4.09232996892928E-015</td>
<td>4.09232996892928E-015</td>
</tr>
</tbody>
</table>
Henrici’s transformation

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

<table>
<thead>
<tr>
<th>iter</th>
<th>GMO</th>
<th>MTH</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2.828427124746190E-000</td>
<td>2.828427124746190E-000</td>
</tr>
<tr>
<td>1</td>
<td>3.999999999999999E-001</td>
<td>2.346678418411335E-002</td>
</tr>
<tr>
<td>2</td>
<td>1.154829390669024E-001</td>
<td>2.555952658965605E-003</td>
</tr>
<tr>
<td>4</td>
<td>3.535120882045415E-003</td>
<td>5.075870513349970E-006</td>
</tr>
<tr>
<td>6</td>
<td>1.125058682735095E-004</td>
<td>5.373992891221222E-009</td>
</tr>
<tr>
<td>8</td>
<td>3.584010172629917E-006</td>
<td>5.461683032157274E-012</td>
</tr>
<tr>
<td>10</td>
<td>1.141764851087469E-004</td>
<td>5.543210949557351E-015</td>
</tr>
<tr>
<td>12</td>
<td>3.637345311678336E-009</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>2.068216140766206E-011</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>3.821435084792773E-015</td>
<td></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>iter</th>
<th>GMO</th>
<th>MTH</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5.656854249492379E-001</td>
<td>5.656854249492379E-001</td>
</tr>
<tr>
<td>1</td>
<td>1.374420072475986E-001</td>
<td>1.476869385644176E-003</td>
</tr>
<tr>
<td>2</td>
<td>6.44864012207421E-003</td>
<td>3.107703683385792E-005</td>
</tr>
<tr>
<td>4</td>
<td>2.982170657758973E-005</td>
<td>4.235319264903191E-010</td>
</tr>
<tr>
<td>6</td>
<td>1.333807714493249E-007</td>
<td>8.461361619985638E-015</td>
</tr>
<tr>
<td>8</td>
<td>5.964656961345872E-010</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>2.667220740432435E-012</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>8.437427167243345E-016</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 1. Example 5 with $n = 2$, initial point = $(-1, 2)$
Example 5. This example is taken from [5]. We consider the function
\[ f(x) = \sum_{1 \leq i \leq n} f_i^2(x), \]
where \( n \) is any positive multiple of 2, \( x = (x_i)_{1 \leq i \leq n} \), 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


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