TRUST REGION WITH NONLINEAR CONJUGATE GRADIENT METHOD

Dalal Hachelfi\textsuperscript{1}, Yamina Laskri\textsuperscript{2}, Mohamed Lamine Sahari\textsuperscript{3}

\textsuperscript{1,2,3}Department of Mathematics
Badji Mokhtar University
P.O. Box 12, Annaba, 23000, ALGERIA
\textsuperscript{1}dalelhach@yahoo.fr

Abstract: Descent direction methods and trust region methods are usually used to solve the unconstrained optimization problem \((p)\)

\[ \min_{x \in \mathbb{R}^n} f(x). \]

In this work, we are interested in convergence results that use trust region methods which employ the conjugate gradient method Day-Yuan version as a subprogram for each iteration. Further, we penalize the quadratic problems with constraints and convert them into series of unconstrained problems.

AMS Subject Classification: 65K05, 90C30

Key Words: unconstrained optimization, conjugate gradient method Day-Yuan version, Wolf’s rule, trust region methods, penalty method

1. Introduction

The aim of this paper is to solve the following unconstrained optimization problem

\[ (P): \min_{x \in \mathbb{R}^n} f(x), \tag{1.1} \]

where the function \(f\) is assumed to be nonlinear and differentiable.
At the current point $x_k \in \mathbb{R}^n$, a model of the variation of $f$, for an increment $d$ of $x_k$, is supposed to be given. In differentiable optimization, it is reasonable to consider a quadratic model of the form

$$f(x_k + d) \simeq \Psi_k(d) \simeq f(x_k) + g_k^T d + \frac{1}{2} d^T H_k d,$$

where $g_k = \nabla f(x_k)$ is the gradient of $f$ at $x_k$ and $H_k = \nabla^2 f(x_k)$ is the Hessian of $f$ at $x_k$. $g_k$ and $H_k$ are assumed to be computed using the Euclidean scalar product $\langle u, v \rangle = \sum_i u_i v_i$.

In trust region methods\cite{1, 2, 3, 7}, we consider that $\Psi_k$ is a model of the variation of $f$ which is acceptable in a neighborhood of the form

$$\overline{B}(0, \triangle_k) = \{ x \in \mathbb{R}^n : \|x\| \leq \triangle_k \},$$

where $\triangle_k > 0$ and $\|\cdot\|$ is the Euclidean norm. The domain $\overline{B}(0, \triangle_k)$ is called the **Trust region** of the model $\Psi_k$ and the positive number $\triangle_k$ is called the **radius** of trust.

To find the increase $d_k$ to be given to $x_k$, we minimize the quadratic model $\Psi_k$ on the trust region \cite{7}. Therefore, we have to solve the quadratic subproblem

$$(p_1) \left\{ \begin{array}{l} \min \Psi_k(d) \\ \|d\| \leq \triangle_k \end{array} \right.$$

A possible approach for solving constraint problems is the **penalty method** \cite{7}. Clearly, we introduce the constraints in the objective under penalties form by starting from a non-feasible solution and try to impose on the unconstrained optimum to arrive to qualifying set.

By solving a problem of the form

$$(p_2) \left\{ \begin{array}{l} \min \Phi(d, \mu_k) = \min_{d \in \mathbb{R}^n} \{ \Psi(d) + \mu_k p(d) \} \end{array} \right.$$

the sequence of values $\{ \mu_k \}$ has the following properties

i) $\mu_k > 0$, $\forall k$,

ii) the sequence $\{ \mu_k \}$ is increasing. That is $\mu_{k+1} > \mu_k$ for every $k$,

iii) the sequence $\{ \mu_k \}$ diverges as $k \to \infty$,\[\text{Page } 300\]
and the $p(d)$ is a continuous penalty function such that

$$p(d) = \begin{cases} 0 & \text{if } \|d\| \leq \Delta \\ > 0 & \text{otherwise.} \end{cases} \quad (1.5)$$

We choose

$$p(d) = (\max\{0, \|d\| - \Delta\})^2$$

2. Algorithm 1.

Initialization step. Set constants $\eta_1, \eta_2, \gamma_1, \gamma_2$ and $\gamma_3$ such as

$$0 < \eta_1 < \eta_2 < 1 \text{ et } 0 < \gamma_1 < \gamma_2 < 1 \leq \gamma_3. \quad (2.1)$$

Set the trust radius $\Delta_0$ and an iterate $x_0$. Compute of $f(x_0)$ . Put $k = 0$ and go to step 1.

Step 1. Construct the model function $\Phi(d, \mu_k)$ according to the relation (1.4) and determine an approximate solution $d_k$ of the problem $(p_2)$

$$(p_2) \left\{ \min_{d \in \mathbb{R}^n} \Phi(d, \mu_k) \right\}.$$

Step 1.0. Set a penalized problem $\phi_k(d, \mu_k)$ and $\mu_0 > 0$, $\beta > 1$, the start point $s_0$, $M_0 = \nabla \phi(d_0, \mu_0)$, put $s_0 = -M_0$. Set $k = 0$ and go to step 1.1.

Step 1.1. If $M_k = 0$: stop ($d^* = d_k$). “Stop test”. Otherwise go to step 1.2.

Step 1.2. Put $\mu_{k+1} = \beta \mu_k$.
Define $d_{k+1} = d_k + \alpha_k s_k$ with $\alpha_k$ satisfies the conditions of low Wolfe

$$\left\{ \begin{array}{l} \phi(d_k + \alpha_k s_k, \mu_k) \leq \phi(d_k, \mu_k) + \rho \alpha_k s_k^T M_k \\ s_k^T M_{k+1} \geq \sigma s_k^T M_k \end{array} \right. \text{ such that } 0 < \rho < \sigma < 1.$$

$s_k = -M_{k+1} + B_{k+1} s_k$
\[ B^{DY}_{k+1} = \frac{\|M_{k+1}\|^2}{s_k^T [M_{k+1} - M_k]} = \frac{\|M_{k+1}\|^2}{s_k^T y_k}. \]

Set \( k = k + 1 \) and go to step 1.1.

**Step 2.** Compute of \( f(x_k + d_k) \) and evaluate the performance criteria

\[ \rho_k = \frac{f(x_k) - f(x_k + d_k)}{\Psi(x_k) - \Psi(x_k + d_k)}. \]

**Step 3.** Update the radius of trust

(a) If \( \rho_k \geq \eta_2 \): The step is a success and \( d_k \) is accepted. \( x_{k+1} = x_k + d_k \) and choose \( \Delta_k \in [\Delta_k, \gamma_3 \Delta_k] \).

(b) If \( \eta_1 \leq \rho_k \leq \eta_2 \): The step is a success and \( d_k \) is accepted. \( x_{k+1} = x_k + d_k \) and choose \( \Delta_k \in [\gamma_2 \Delta_k, \Delta_k] \).

(c) If \( \rho_k < \eta_1 \): The step is a failure and \( d_k \) is rejected. \( x_{k+1} = x_k \) and choose \( \Delta_k \in [\gamma_1 \Delta_k, \gamma_2 \Delta_k] \).

**Step 4.** If \( \|\nabla f(x_k)\| \leq \varepsilon \), then the algorithm is stopped. Otherwise, we set \( k = k + 1 \) and go back to step 1.

We usually choose \( \varepsilon = 10^{-5}, \eta_1 = 0.25, \eta_2 = 0.75, \gamma_1 = 0.25, \gamma_2 = 0.75 \) and \( \gamma_3 = 2 \).

### 3. Convergence Results

By a similar manner to the linear search methods, the determination of optimal steps \( d_k \) is not a necessary condition for global convergence. Under certain conditions, a good approximation of these steps may be acceptable. It suffices to determine an approximate solution \( d_k \), in the interior of the trust region, which produces a sufficient reduction of the model function. This reduction can be achieved by the method of Cauchy point \( d_k^c \).

**Definition 1.** [3] We call Cauchy point of quadratic subproblem \( (RC_k) \), the point noted \( d_k^c \) solution of the problem

\[
(p_3) \begin{cases} 
\min \Psi_k(d) \\
\|d\| \leq \Delta_k \\
\quad d = -\alpha g_k, \alpha \in \mathbb{R}.
\end{cases}
\]
Therefore, such a point is the point minimizing $\Psi_k$ in the trust region. That is, along the right side of the strongest slope of $\Psi_k$.

**Proposition 2.** [3] The Cauchy point $d^c_k$ is unique and it is given by

$$
d^c_k = \begin{cases} 
0 & \text{if } g_k = 0 \\
-\frac{\triangle_k}{\|g_k\|} g_k & \text{if } g_k \neq 0 \text{ et } g_k^T H_k g_k \leq 0 \\
-\frac{\triangle_k}{\|g_k\|} g_k \min \left( \frac{\|g_k\|^3}{\triangle_k g_k^T H_k g_k}, 1 \right) & \text{otherwise}.
\end{cases}
$$

\[(3.1)\]

**Proof.** Indeed, the case where $g_k = 0$ is obvious. Suppose now that $g_k \neq 0$. If $g_k^T H_k g_k \leq 0$, then $\alpha \mapsto \Psi_k(-\alpha g_k)$ is concave and $\nabla \Psi_k(0)^T (-g_k) = -\|g_k\|^2 < 0$, therefore, $\alpha$ must be taken as large as possible while keeping $\|\alpha g_k\| \leq \triangle_k$. This gives $\alpha = \frac{\triangle_k}{\|g_k\|}$.

It remains to examine the case $g_k^T H_k g_k > 0$. The result follows from the fact that, in this case, the minimum of the function $\alpha \mapsto \Psi_k(-\alpha g_k)$ is attained for $\alpha = \frac{\|g_k\|^2}{g_k^T H_k g_k}$.

As we have seen, the point of Cauchy $d^c_k$ allows whether an approximate solution of the problem in trust region to be validated or not. So, the point $d_k$ must verify the decrease of the model function

$$
\Psi_k(0) - \Psi_k(d_k) \geq c_1 (\Psi_k(0) - \Psi_k(d^c_k)). \quad c_1 > 0
$$

\[(3.2)\]

The following proposition is known as Powell Condition.

**Proposition 3.** [1] The Cauchy point $d^c_k$ verifies the condition of Powell with $C = \frac{1}{2}$, such as

$$
\Psi_k(0) - \Psi_k(d^c_k) \geq C \|g_k\| \min \left\{ \frac{\|g_k\|}{\|H_k\|}, \triangle_k \right\}.
$$

\[(3.3)\]

**Proof.** We consider two cases for the Hessian matrix $H_k$. That is,

1. $g_k^T H_k g_k \leq 0$

$$
\Psi_k(d^c_k) - \Psi_k(0) = \Psi_k \left( -\frac{\triangle_k}{\|g_k\|} g_k \right) - f_k
$$
\[ \begin{align*}
&= -\frac{\triangle_k}{\|g_k\|} \|g_k\|^2 + \frac{1}{2} \frac{\triangle_k^2}{\|g_k\|^2} g_k^T H_k g_k \\
&\leq -\frac{\triangle_k}{\|g_k\|} \|g_k\|^2 \\
&\leq -\triangle_k \|g_k\| \\
&\leq -\|g_k\| \min \left\{ \frac{\|g_k\|}{\|H_k\|}, \triangle_k \right\} \\
&\leq -\frac{1}{2} \|g_k\| \min \left\{ \frac{\|g_k\|}{\|H_k\|}, \triangle_k \right\}
\end{align*} \]

so, we have (3.3).

b) \( g_k^T H_k g_k > 0 \) and \( \frac{\|g_k\|^3}{\triangle_k g_k^T H_k g_k} \leq 1: \)

\[ \begin{align*}
\Psi_k(d_k^c) - \Psi_k(0) &= -\frac{\|g_k\|^4}{\triangle_k g_k^T H_k g_k} + \frac{1}{2} g_k^T H_k g_k \frac{\|g_k\|^4}{(g_k^T H_k g_k)^2} \\
&= -\frac{1}{2} \frac{\|g_k\|^4}{g_k^T H_k g_k} \\
&\leq -\frac{1}{2} \frac{\|g_k\|^4}{\|H_k\| \|g_k\|^2} \\
&\leq -\frac{1}{2} \frac{\|g_k\|^2}{\|H_k\|} \\
&\leq -\frac{1}{2} \|g_k\| \min \left\{ \frac{\|g_k\|}{\|H_k\|}, \triangle_k \right\}
\end{align*} \]

so, we have (3.3).

c) \( g_k^T H_k g_k < \frac{\|g_k\|^3}{\triangle_k}, \ d_k^c = -\frac{\triangle_k}{\|g_k\|} g_k \)

we have (3.3).

\( \square \)

**Theorem 4.** [1] Let \( d_k \) be an arbitrary vector such that

\[ \|d_k\| \leq \triangle_k \]

and

\[ \Psi_k(0) - \Psi_k(d_k) \geq c_2 (\Psi_k(0) - \Psi_k(d_k^c)). \]  \( (3.4) \)
Then, $d_k$ satisfies (3.3) with $c_2 = \frac{c_1}{2}$.

Proof. Since $\|d_k\| \leq \Delta_k$, from (3.2), we have,

$$
\Psi_k(0) - \Psi_k(d_k) \geq c_1 (\Psi_k(0) - \Psi_k(d_k^C)) \geq \frac{1}{2} c_1 \parallel g_k \parallel \min \left\{ \parallel g_k \parallel | H_k \parallel, \Delta_k \right\}
$$

In particular, if $d_k$ is the exact solution of $(RC_k)$, then it satisfies (3.3) with $c_2 = \frac{1}{2}$.

3.1. Convergence of penalty methods

The following two lemmas are used to prove the convergence of the method.

By $d_k$ we denote the solution of $(p_2)$.

Lemma 5. [7] For any value of $k$, we have

i) $\Phi (d_k, \mu_k) \leq \Phi (d_{k+1}, \mu_{k+1})$

ii) $p(d_k) \geq p(d_{k+1})$

iii) $\Psi(d_k) \leq \Psi(d_{k+1})$

Proof. •

i) Since the sequence $\{\mu_k\}$ is increasing and $p(d_{k+1}) \geq 0$, we have

$$
\Phi (d_{k+1}, \mu_{k+1}) = \Psi(d_{k+1}) + \mu_{k+1} p(d_{k+1}) \\
\geq \Psi(d_{k+1}) + \mu_k p(d_{k+1}) \\
\geq \Psi(d_k) + \mu_k p(d_k)
$$

and $d_k$ minimizes $\Phi (d, \mu_k)$, we obtain

$$
\Phi (d_{k+1}, \mu_{k+1}) \geq \Phi (d_k, \mu_k).
$$

ii) Since $d_k$ minimizes $\Phi (d, \mu_k)$

$$
\Phi (d, \mu_k) \Psi(d_k) + \mu_k p(d_k) \leq \Psi(d_{k+1}) + \mu_k p(d_{k+1})
$$

and $d_{k+1}$ minimizes $\Phi (d, \mu_{k+1})$

$$
\Psi(d_{k+1}) + \mu_{k+1} p(d_{k+1}) \leq \Psi(d_k) + \mu_{k+1} p(d_k),
$$
adding the two inequalities we obtain,

\[(\mu_{k+1} - \mu_k)p(d_{k+1}) \leq (\mu_{k+1} - \mu_k)p(d_k).
\]

Since \(\mu_{k+1} > \mu_k\), we have

\[p(d_{k+1}) \leq p(d_k).
\]

iii) \(\Psi(d_{k+1}) + \mu_k p(d_{k+1}) \geq \Psi(d_k) + \mu_k p(d_k)\) using ii) and the fact that \(\mu_k \geq 0\), we have

\[\Psi(d_{k+1}) + \mu_k p(d_{k+1}) \geq \Psi(d_k) + \mu_k p(d_k+1).
\]

Consequently, \(\Psi(d_{k+1}) \geq \Psi(d_k)\).

Lemma 6. [7] \(d^*\) be an optimal solution of \((p_1)\). Then, for each \(k\) we have,

\[\Psi(d^*) \geq \Phi(d_k, \mu_k) \geq \Psi(d_k).
\]

Proof.

\[
\Psi(d^*) = \Psi(d^*) + \mu_k p(d^*) \forall k, \text{ since } p(d^*) = 0
\]

\[\geq \Psi(d_k) + \mu_k p(d_k) = \Phi(d_k, \mu_k) \forall k \text{ (since } d_k \text{ minimizes } \Phi(d, \mu_k))
\]

\[\geq \Psi(d_k) \forall k \text{ (since } \mu_k \geq 0 \text{ and } p(d_k) \geq 0).\]

So, \(\Psi(d^*) \geq \Phi(d_k, \mu_k) \geq \Psi(d_k) \forall k\). \qed

Theorem 7. Suppose that \(d_k\) is an exact global minimum of \((p_2)\) and assume that \(\{\mu_k\}\) is a divergent sequence. Then, every limit point \(d^*\) of the sequence \(\{d_k\}\) is a solution of \((p_1)\).

Proof. Let \(\overline{d}\) be the solution of \((p_1)\) such that

\[\Psi(\overline{d}) \leq \Psi(d) \forall \|d\| \leq \Delta.
\]

Since \(d_k\) minimizes \(\Phi(d, \mu_k)\) for all \(k\), we have,

\[\Phi(d_k, \mu_k) \leq \Phi(\overline{d}, \mu_k).
\]

So,

\[\Psi(d_k) + \mu_k p(d_k) \leq \Psi(\overline{d}) + \mu_k p(\overline{d}) = \Psi(\overline{d}) \text{ (because } p(\overline{d}) = 0).\]
Then,
\[ p(d_k) \leq \frac{1}{\mu_k} [\Psi(d) - \Psi(d_k)] \]  
(3.5)

It is assumed that \( d^* \) is the limit point \( \{d_k\} \). Then, there exists an infinite subsequence of \( K \) such that
\[ \lim_{k \in K} d_k = d^*. \]

By taking the limit of inequality (3.5), we find
\[ p(d^*) = \lim_{k \in K} p(d_k) \leq \lim_{k \in K} \frac{1}{\mu_k} [\Psi(d) - \Psi(d_k)] = 0 \]
(because \( \{\mu_k\} \) is a divergent sequence).

So, we have \( p(d^*) = 0 \) and consequently \( \|d^*\| \leq \triangle \).

The optimality of \( d^* \) follows directly from lemma 6. Indeed, the relation \( \Psi(d_k) \leq \Psi(d^*) \) being satisfied for any \( k \), it follows that
\[ \Psi(d^*) = \lim_{k \in K} \Psi(d_k) \leq \Psi(d). \]

\[ \square \]

3.2. Global Convergence

**Theorem 8.** Let \( f : \mathbb{R}^n \to \mathbb{R} \) be a continuously differentiable function on the set
\[ X = \{x \in \mathbb{R}^n : f(x) < f(x_0)\}, \]

and let \( \{x_k\} \) be a sequence generated by the algorithm 1, with \( H_k \) uniformly bounded. If \( d_k \) satisfies the condition (3.2) then we have
\[ \liminf_{k \to \infty} \|\nabla f_k\| = 0. \]  
(3.6)

**Proof.** Assume that there exists an indice \( k_1 \) and a constant \( \gamma > 0 \) such that
\[ \|\nabla f_k\| \geq \gamma, \forall k \geq k_1. \]  
(3.7)

We claim that
\[ \triangle_k \to 0 \quad \text{and} \quad \sum_{k \geq 1} \|d_k\| < \infty. \]  
(3.8)
According to the condition of Powell (3.3) and the fact that \( \{H_k\} \) is bounded \([3]\), we have

\[
\Psi_k(d_k) \leq -C \| \nabla f_k \| \min \left\{ \frac{\| \nabla f_k \|}{\| H_k \|}, \triangle_k \right\} \leq -C \min \{1, \triangle_k\},
\]

where \( C \) is a positive constant which is independent of \( k \). Using the step 3 of the algorithm 1 (\( \rho_k \geq \eta_2 \) ), we have

\[
f(x_k + d_k) \leq f(x_k) + \eta_2 \Psi_k(d_k).
\]

The last inequality can be written as

\[
f(x_{k+1}) \leq f(x_k) - C \min \{1, \triangle_k\},
\]

\[
C \min \{1, \triangle_k\} \leq f(x_k) - f(x_{k+1}).
\]

Since \( f \) is bounded below, the sequence \( \{f(x_k) - f(x_{k+1})\} \) is convergent and we deduce from the above inequality that \( \sum_{k \geq 1} \triangle_k < \infty \), and since \( \|d_k\| \leq \triangle_k \), the assertions of (3.8) are deduced.

We now show that the ratio \( \rho_k \to 1 \). Since \( f \) is bounded and \( \|d_k\| \leq \triangle_k \) then,

\[
|f(x_k + d_k) - f(x_k) - \nabla f(x_k)^T d_k| \leq C \|d_k\|^2.
\]

On the other hand, since \( \{H_k\} \) is bounded, we have:

\[
|\Psi_k(d_k) - \nabla f(x_k)^T d_k| \leq c \|d_k\|^2.
\]

Finally, we can successively write

\[
\rho_k = \frac{f(x_{k+1}) - f(x_k)}{\Psi_k(d_k)} = \frac{f(x_k + d_k) - f(x_k) - \nabla f(x_k)^T d_k}{\Psi_k(d_k)} + \frac{\nabla f(x_k)^T d_k - \Psi_k(d_k)}{\Psi_k(d_k)} + 1
\]

\[
|\rho_k - 1| \leq \frac{f(x_k + d_k) - f(x_k) - \nabla f(x_k)^T d_k}{|\Psi_k(d_k)|}.
\]
\[ + \frac{\nabla f(x_k)^T d_k - \Psi_k(d_k)}{|\Psi_k(d_k)|}. \]

From (3.9), \( \Delta \to 0 \) and \( \|s_k\| \leq \Delta_k \), and so,
\[ |\Psi_k(d_k)| \geq c \|d_k\|. \]
Then, combining (3.9) and (3.10), we get
\[ |\rho_k - 1| \leq \frac{c \|d_k\|^2}{c \|d_k\|} \leq \|d_k\|. \]
This shows that \( \rho_k \to 1 \).

Accordingly, \( \rho_k > \eta_1 \) for \( k \) large enough. Due to the updating rule of \( \Delta_k \) (see the step 3 of the algorithm 1), this implies that \( \Delta_k > \Delta > 0 \). But this contradicts the first assertion of (3.8). This contradiction completes the proof.

**Theorem 9.** Suppose, in addition to the assumptions and conditions of Theorem 8, that \( \nabla f_k \) is continuous in the sense of Lipschitz and bounded below on the set
\[ X = \{ x \in \mathbb{R}^n : f(x) < f(x_0) \}, \]
then we have
\[ \lim_{k \to \infty} \|\nabla f_k\| = 0. \quad (3.11) \]

**Proof.** Suppose that there exists a constant \( \varepsilon > 0 \) and a subsequence \( \{x_k\}_{k \geq m} \subset B(x_m, R) \) where \( B(x_m, R) \) is the closed ball of center \( x_m \) and radius \( R \), for which we have
\[ \|\nabla f_k\| \geq \varepsilon, \quad \text{for all } k \geq m. \quad (3.12) \]
Let \( l \geq m \) such that \( x_{k+1} \) is the first iterate after \( x_m \) outside \( B(x_m, R) \). Since \( \|\nabla f_k\| \geq \varepsilon \) for all \( k = m, m + 1, \ldots, l \), we can write
\[ f(x_m) - f(x_{l+1}) = \sum_{k=m}^{l} f(x_k) - f(x_{k+1}). \]

Using the condition of Powell (3.3) and the fact that \( \{H_k\} \) is bounded [3], we have as in the proof of Theorem 8,
\[ f(x_{k+1}) - f(x_k) \leq -C \min \{ \|\nabla f_k\|, \triangle_k \}. \]

So, for all \( k \geq m \), we have

\[ f(x_{l+1}) - f(x_m) \leq -C \min \{1, \triangle_k\}. \]

Since \( f(x_{l+1}) - f(x_m) \to 0 \), \( \min \{1, \triangle_k\} = \triangle_k \), for \( k \) large enough [3]. Finally, using the fact that \( \|d_k\| \leq \triangle_k \), we have

\[ \|d_k\| \leq C (f(x_k) - f(x_{l+1})) , \quad k = m, m + 1, \ldots, l. \]

\[ \|x_{l+1} - x_m\| \leq \sum_{k=m}^{l} \|d_k\| \leq C (f(x_m) - f(x_{l+1})). \]

So \( \|x_{l+1} - x_m\| \to 0 \). Then by the uniform continuity of \( \nabla f_k \), we have

\[ \|f(x_{l+1}) - f(x_m)\| \to 0. \]

This is in contradiction with (3.12). Hence the theorem is proved. \( \square \)

4. Numerical tests

We have performed numerical tests using the trust region algorithm penalized by the following functions

Problem 1: The generalized function of Rosenbrock

\[ f(x) = \sum_{i=1}^{n-1} 100(x_i^2 - x_{i+1})^2 + (x_i - 1)^2 \]

Problem 2: The generalized function of Dixon

\[ f(x) = (x_1 - 1)^2 + \sum_{i=2}^{n} i(2x_i^2 - x_{i-1})^2 \]

Problem 3: The function of Oren

\[ f(x) = \left[ \sum_{i=1}^{n} ix_i^2 \right]^2 \]
Problem 4: The function of Powell

\[
f(x) = \sum_{i=1}^{n/4} \left[ (x_{4i-3} + 10x_{4i-2})^2 + 5(x_{4i-1} - x_{4i})^2 
+ (x_{4i-2} - 2x_{4i-1})^4 + 10(x_{4i-3} - x_{4i})^4 \right]
\]

In all tests the stopping criterion of the algorithm is \( \varepsilon = 10^{-5} \). The trust regions algorithm is applied with \( \Delta_0 = 0.5, \eta_1 = 0.25, \eta_2 = 0.75, \gamma_1 = 0.25, \gamma_2 = 0.75 \) and \( \gamma_3 = 2 \). The conjugate gradient algorithm is applied to the linear search of wolfe with the parameters: \( m_1 = 0.3, m_2 = 0.7 \). The penalty method is applied with \( \mu_0 = 2 \).

The results showing the number of iterations and time of convergence are illustrated in the following tables.

### Table 1: Rosenbrock function

<table>
<thead>
<tr>
<th>Dimension n</th>
<th>iteration k</th>
<th>( f(x_k) )</th>
<th>( |g(x_k)| )</th>
<th>Time of convergence</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>26</td>
<td>2, 59.10^{-12}</td>
<td>9, 56.10^{-6}</td>
<td>4, 68.10^{-2} s</td>
</tr>
<tr>
<td>10</td>
<td>29</td>
<td>4, 82.10^{-11}</td>
<td>8, 71.10^{-6}</td>
<td>0, 26 s</td>
</tr>
<tr>
<td>30</td>
<td>29</td>
<td>8, 16.10^{-13}</td>
<td>1, 27.10^{-6}</td>
<td>2, 51 s</td>
</tr>
<tr>
<td>50</td>
<td>28</td>
<td>2, 72.10^{-11}</td>
<td>7, 91.10^{-6}</td>
<td>5, 47 s</td>
</tr>
<tr>
<td>80</td>
<td>30</td>
<td>3.10^{-11}</td>
<td>6, 46.10^{-6}</td>
<td>20, 57 s</td>
</tr>
<tr>
<td>100</td>
<td>28</td>
<td>2, 71.10^{-11}</td>
<td>8, 59.10^{-6}</td>
<td>22 s</td>
</tr>
<tr>
<td>150</td>
<td>27</td>
<td>5, 67.10^{-12}</td>
<td>6, 53.10^{-6}</td>
<td>50, 66 s</td>
</tr>
<tr>
<td>200</td>
<td>25</td>
<td>2, 11.10^{-13}</td>
<td>5, 46.10^{-6}</td>
<td>129, 62 s</td>
</tr>
</tbody>
</table>

### Table 2: Dixon function

<table>
<thead>
<tr>
<th>Dimension n</th>
<th>iteration k</th>
<th>( f(x_k) )</th>
<th>( |g(x_k)| )</th>
<th>Time of convergence</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>8</td>
<td>5, 62.10^{-12}</td>
<td>8, 72.10^{-6}</td>
<td>1, 56.10^{-2} s</td>
</tr>
<tr>
<td>6</td>
<td>25</td>
<td>1, 90.10^{-12}</td>
<td>9, 97.10^{-6}</td>
<td>1, 06 s</td>
</tr>
<tr>
<td>10</td>
<td>41</td>
<td>0, 66</td>
<td>8, 43.10^{-6}</td>
<td>1, 29 s</td>
</tr>
<tr>
<td>30</td>
<td>68</td>
<td>0, 66</td>
<td>9, 79.10^{-6}</td>
<td>2, 72 s</td>
</tr>
<tr>
<td>50</td>
<td>111</td>
<td>0, 66</td>
<td>9, 45.10^{-6}</td>
<td>6, 83 s</td>
</tr>
<tr>
<td>100</td>
<td>36</td>
<td>0, 66</td>
<td>8, 84.10^{-6}</td>
<td>39, 18 s</td>
</tr>
</tbody>
</table>
Table 3: Oren Function

<table>
<thead>
<tr>
<th>Dimension n</th>
<th>iteration k</th>
<th>$f(x_k)$</th>
<th>$|g(x_k)|$</th>
<th>Time of convergence</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>7</td>
<td>$2.68.10^{-8}$</td>
<td>$9.84.10^{-6}$</td>
<td>$1.56.10^{-2}$ s</td>
</tr>
<tr>
<td>10</td>
<td>19</td>
<td>$1.10.10^{-8}$</td>
<td>$9.66.10^{-6}$</td>
<td>0.60 s</td>
</tr>
<tr>
<td>50</td>
<td>23</td>
<td>$4.20.10^{-9}$</td>
<td>$9.74.10^{-6}$</td>
<td>103, 13 s</td>
</tr>
<tr>
<td>80</td>
<td>25</td>
<td>$3.13.10^{-8}$</td>
<td>$9.94.10^{-6}$</td>
<td>470 s</td>
</tr>
<tr>
<td>100</td>
<td>25</td>
<td>$2.52.10^{-9}$</td>
<td>$9.32.10^{-6}$</td>
<td>967 s</td>
</tr>
</tbody>
</table>

Table 4: Powell function

<table>
<thead>
<tr>
<th>Dimension n</th>
<th>iteration k</th>
<th>$f(x_k)$</th>
<th>$|g(x_k)|$</th>
<th>Time of convergence</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>20</td>
<td>$1.15.10^{-8}$</td>
<td>$9.90.10^{-6}$</td>
<td>0.15 s</td>
</tr>
<tr>
<td>8</td>
<td>53</td>
<td>$1.76.10^{-8}$</td>
<td>$8.89.10^{-6}$</td>
<td>0.24 s</td>
</tr>
<tr>
<td>16</td>
<td>79</td>
<td>$9.38.10^{-9}$</td>
<td>$9.28.10^{-6}$</td>
<td>1.04 s</td>
</tr>
<tr>
<td>32</td>
<td>18</td>
<td>$1.25.10^{-8}$</td>
<td>$4.40.10^{-6}$</td>
<td>2.72 s</td>
</tr>
<tr>
<td>100</td>
<td>19</td>
<td>$2.02.10^{-8}$</td>
<td>$9.30.10^{-6}$</td>
<td>31, 90 s</td>
</tr>
<tr>
<td>200</td>
<td>20</td>
<td>$3.86.10^{-8}$</td>
<td>$5.92.10^{-6}$</td>
<td>138, 37 s</td>
</tr>
</tbody>
</table>

References


TRUST REGION WITH NONLINEAR CONJUGATE...


